Journal of Experimental Psychology:
Human Perception and Performance
1987, Vol. 13, No. 3, 335-343

0096-1523/87/$00.75

Minimum Points and Views for the Recovery of Three-Dimensional Structure

Myron L. Braunstein, Donald D. Hoffman, and Lionel R. Shapiro
University of California, Irvine

George J. Andersen
University of Illinois at Urbana-Champaign

Bruce M. Bennett
Department of Mathematics
University of California, Irvine

Mathematical analyses of motion perception have established minimum combinations of points and distinct views that are sufficient to recover three-dimensional (3D) structure from two-dimensional (2D) images, using such regularities as rigid motion, fixed axis of rotation, and constant angular velocity. To determine whether human subjects could recover 3D information at these theoretical levels, we presented subjects with pairs of displays and asked them to determine whether they represented the same or different 3D structures. Number of points was varied between two and five; number of views was varied between two and six; and the motion was fixed axis with constant angular velocity, fixed axis with variable velocity, or variable axis with variable velocity. Accuracy increased with views, decreased with points, and was greater with fixed-axis motion. Subjects performed above chance levels even when motion was eliminated, indicating that they exploited regularities in addition to those in the theoretical analyses.

Theoretical investigations of visual motion have provided a number of specific analyses of the minimum number of points and views required to recover three-dimensional (3D) structure from two-dimensional (2D) images. Recovery of 3D structure, in this context, is defined as determining the x, y, and z coordinates of each point, up to a scale factor. These analyses differ in the constraints that are imposed. Ullman (1979) showed that under a rigidity constraint, three views of four noncoplanar points are sufficient to recover structure in an orthographic projection, up to a reflection about the frontal plane. The required numbers of points and views are reduced by adding further constraints, such as planarity (Hoffman & Flinchbaugh, 1982), fixed axis of rotation (Hoffman & Bennett, 1986; Webb & Aggarwal, 1981), and constant angular velocity (Hoffman & Bennett, 1985). These proofs are summarized in Table 1.

A number of empirical studies have addressed issues related to theoretical analyses of the recovery of structure from motion. Several studies (e.g., Braunstein & Andersen, 1986; Schwartz & Sperling, 1983; Todd, 1985) have questioned the generality of the rigidity constraint. Other studies have considered the recovery of structure with small numbers of views or with small numbers of points. Lappin, Doner, and Kottas (1980) found that subjects could make accurate judgments based on 3D structure with two perspective views of 512 points. Lappin and Fuqua (1983) found a high level of accuracy for relative depth judgments with 120° rotations of three-point configurations. There have not been studies, however, of the recovery of structure from motion using the minimum combinations of points and views found in the theoretical analyses discussed above.

There are several reasons why these theoretical analyses should be considered empirically. First, it is worthwhile to determine whether the performance of human observers approaches the performance of the "ideal" observers in these analyses. Can the human observer recover 3D structure at the minimum combinations of points and views? Second, it is useful to know whether performance improves as predicted by these theoretical analyses when constraints in addition to that of rigidity are imposed on the displays. Specifically, can structure be recovered with fewer points and views when the axis of rotation is fixed and when a constant angular velocity is maintained across views? Third, empirical studies may suggest other constraints used by human observers that have not been considered in theoretical analyses.

On the negative side, one can question the ecological validity of minimum information displays and of orthographic projections in particular. These displays are clearly special cases. Visual perception normally occurs in richly textured environments with continuous observation. Orthographic projection simulates an infinite viewing distance, eliminating the perspective effects found in normal vision. With these considerations in mind, we still believe that these displays provide a useful starting point for bringing together specific mathematical analyses with psychophysical procedures.

There are at least two fundamental difficulties in applying a psychophysical approach to the testing of theoretical analyses

This research was supported by a contract to D. Hoffman from the Office of Naval Research, Cognitive and Neural Sciences Division, Perceptual Sciences Group. We thank Joseph Lappin and James Todd for helpful comments on an earlier version of this article and Johnna Eastburn and James Tittle for assistance in various aspects of this research.

Correspondence concerning this article should be addressed to Myron L. Braunstein, Cognitive Sciences Group, School of Social Sciences, University of California, Irvine, California 92717.

Table 1
Sufficient Conditions for the Recovery of Three-Dimensional Structure

Number of distinct views	Number of points		
	2	3	4
2		Pairwise-rigid and planar motion[a]	
3	Rigid planar motion[a]	Rigid fixed-axis motion[b]	Rigid motion[c]
	Rigid fixed-axis motion parallel to image plane, constant angular velocity[b]		Nonrigid fixed-axis motion[d]
4	Nonrigid fixed-axis motion[d]		
	Rigid fixed-axis motion, constant angular velocity[e]		
5	Rigid fixed-axis motion[e]		

[a] Hoffman & Flinchbaugh, 1982. [b] Hoffman & Bennett, 1986. [c] Ullman, 1979. [d] Bennett & Hoffman, 1985. [e] Hoffman & Bennett, 1985.

of the recovery of structure from motion. The first stems from the definition given above, according to which recovery of structure consists of determining coordinates in 3D space. This definition provides a suitable measure for computer simulations, but it is not reasonable to expect a human subject to call out coordinates while observing a group of points undergoing a rotation. Some dependent variable is needed, one that is logically related to the recovery of structure but that is based on a reasonable human response. This will be discussed further in the following paragraphs.

The second difficulty is inherent in the task—recovering 3D structure from 2D images. The information for the recovery of the structure must be available in the images, and therefore any task given the subject to determine whether the structure has been recovered must be possible on the basis of the images. How, then, do we know that the subject is not performing the task on the basis of some 2D characteristic of the images without recovering the 3D structure? There is no way, in principle, to be certain of this. The best we can do is try to find and eliminate any 2D regularities that a subject could use to perform the task without also recovering the 3D structure.

In the first task that we used in pilot studies, displays were generated consisting of three, four, or eight points that were the vertices of regular polygons. Each display was paired with a polygon in which the location of one of the vertices relative to the others was altered by a controlled amount, so that the polygon was no longer regular. The subjects viewed the displays side by side. To prevent direct image comparisons, the two polygons were never displayed in the same orientation. Several experienced subjects reported noticeable regularities in the 2D images of the regular polygons, regularities that made it possible for them to distinguish between the regular and irregular polygons in each pair. Although subjects could not precisely describe all of these apparent regularities, they appeared to be related to the use of regular polygons as the standard stimuli, and we therefore abandoned that approach.

Instead, we used displays consisting of sets of points that were randomly generated (under restrictions described in the Method section). For each display, a comparison display was presented that was identical to the standard or had one point moved to a different position. The subject's task was to indicate whether the 3D structures represented by the two displays were the same or different. The rationale for using a comparison task as a measure of recovery of structure is that subjects must recover the 3D structure in order to determine whether the displays represent the same or different configurations. Although the task can be performed by comparing substructures if the number of points exceeds the minimum required, it should be necessary to use all of the points to recover the 3D structure of a configuration at the minimum levels.

For the reasons stated above, the two displays were presented out of phase. This probably required the subject to mentally rotate one or both structures to compare them. This extra step of mental rotation, between recovery of the structures and the behavioral response, could have prevented subjects from responding accurately (above chance) to the combinations of small numbers of points and small numbers of views listed in Table 1. As we will indicate later, this did not seem to be the case in our experiments.

Each of the mathematical analyses in Table 1 gives sufficiency conditions for recovering the third dimension if one assumes some specific regularity or regularities in the motion of the simulated object. The individual analyses do not make predictions about improvements in performance with increasing numbers of points or views or with further constraints. If all of these analyses were instantiated in the visual system, however, we would expect the following results for the numbers of points and views and the motion constraints included in the present experiment:

1. Accuracy should increase with the number of distinct views.
2. Accuracy should increase with the number of points.
3. Accuracy should increase with increasing constraints, from variable axis to fixed axis to fixed axis with constant angular velocity.

In addition to these general trends, accuracy should increase from chance to above chance at the critical combinations of

points, views, and constraints listed in Table 1: (a) four points, three views, no added constraints (rigidity only); (b) three points, three views, rigidity and fixed-axis constraints; (c) two points, five views, rigidity and fixed-axis constraints; and (d) two points, four views, rigidity, fixed-axis, and constant angular velocity constraints.

Method

Subjects

The subjects were 7 graduate students from the School of Social Sciences at the University of California, Irvine, who were paid for their participation. One subject had taken part in a preliminary experiment; all other subjects were naive. Acuity of at least 20/40 (Snellen eye chart) was required in the eye the subject used throughout the course of the experiment. Each subject met a performance criterion of at least 70% correct judgments during a screening session consisting of 160 trials with 12 views, two to five points, and fixed-axis, constant angular velocity motion. Four of the naive subjects were run without feedback. The remaining subjects were run with feedback on all trials.

Design

We examined four independent variables: the presence or absence of feedback, the number of points in a simulated object, the number of distinct views presented, and motion constraints. The number of points ranged from two to five. The number of distinct views ranged from two to six. Three motion conditions were examined: (a) fixed axis of rotation with constant angular velocity, (b) fixed axis of rotation with a variable angular velocity, and (c) variable axis of rotation. (The motion variable does not strictly apply to the two-view stimuli, but it did affect the selection of views even in that case, determining whether the rotation between views was fixed or selected from a distribution.) All of the independent variables except the feedback variable were run within subjects. Each subject responded to 60 trials in each of the 60 combinations of two-view through six-view conditions and to 60 trials in each of 12 (four numbers of points and three motions) 12-view baseline conditions. In addition, each subject responded to 120 single-view trials at each of the four levels of the point variable.

Stimuli

A stimulus consisted of from two to five light-green dots, changing in position, against a dark-green background. Each stimulus simulated points on a rigid object rotating in depth. Preliminary point positions for an object were selected at random (without replacement) from a uniform distribution of 225 potential point positions on the surface of a unit-radius sphere. To avoid any unintended regularities in the projection that might have resulted from all points being equidistant from the center of rotation, the distance of each point to the center of the sphere was randomly perturbed within a range of ±0.2 units. This configuration of points was defined as the standard object. For *same* trials, the comparison object was identical to the standard object. For *different* trials, the following method was used to generate the comparison object: One of the points on the standard object was moved to one of the 225 potential point positions that was unoccupied. The point to be moved and the new position were selected at random. If the root mean square (RMS) of the changes in distance (standard object distance minus comparison object distance) from the moved point to all other points in the simulated object did not exceed 0.7 units, these simulated objects were discarded and a new standard was generated. The minimum RMS distance criterion was determined, through pilot studies, to provide a better than 0.8 overall proportion correct for an experienced subject with 30-view displays and fixed-axis, constant angular velocity motion. As a result of this criterion, and the restrictions described below, the RMS difference of 3D distances between objects on *different* trials varied between 0.70 and 1.72 (in radius units based on the sphere used to generate the objects), with a mean of 0.94 and a standard deviation of 0.08.

In order to avoid the possibility of subjects' making direct comparisons of the 2D projections, the two simulated objects were set at different initial orientations. The initial slants were varied between 10° and 50°, and initial tilts were varied between 15° and 75°, with a difference of at least 40° in either slant or tilt required between the initial orientations of the standard and comparison objects. (Slant was defined as rotation perpendicular to the image plane; tilt was defined as rotation parallel to the image plane. See Stevens, 1983.) In addition, the standard and comparison objects were always out of phase, with their initial phase difference randomly varied within a range of 40° to 140°.

Each stimulus display consisted of a sequence of orthographic projections (views) of two simulated objects undergoing specific types of motion in three dimensions. In order to allow subjects sufficient time to observe the displays and make a judgment, the sequence of views was oscillated (e.g., $1, 2, \cdots n-1, n, n-1, \cdots 2, 1, 2, \cdots$) at a rate of 16 views per second until the subject responded. (If the subject did not respond within 60 s, the trial was repeated at the end of the session.) For the fixed-axis conditions, these views were rotations from the initial orientation about an axis at 20° slant and 0° tilt for the standard object and at 50° slant and 0° tilt for the comparison object. For constant angular velocity conditions, the rotation between successive views was 6°.

The variability in angular velocity in the variable-velocity condition and the variability in axis of rotation in the variable-axis condition could not be unrestricted. Otherwise, difficulty in maintaining the identity of points from frame to frame (correspondence matches) and in perceiving smooth motion might have confounded the effects of variability. These two factors were controlled first by limiting the variance of the distribution from which the velocities and axis shifts were sampled and then by imposing a correspondence match criterion and a 2D motion criterion (described below) on each display. The axis change and velocity change in the variable-axis and variable-velocity conditions were selected from Gaussian distributions, with means equal to the axis and velocity changes between views in the fixed-axis and fixed-velocity conditions (6°) and standard deviations of 3°. A minimum variance criterion for axis shift or angular rotation was used to control for chance selection of nearly equal axis or rotation values across views in the variable-axis or variable-velocity conditions.

Displays were used only if the following two restrictions were satisfied: In order to reduce the possibility of false correspondence matches of points across pairs of views, the nearest neighbor to any given point, from one view to the next, had to be the correctly corresponding point. This restriction was not applied to points with opposite depth signs, which would be moving in opposite directions. In order to maintain conditions for "short-range" apparent motion across pairs of views, the distance moved by any given point in the image was not allowed to exceed 15′ of visual angle (Braddick, 1974).

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 11/44 computer. The subject viewed the display through a tube arrangement that limited the field of view to a circular area 7.6° in diameter. The maximum projected diameter of each simulated object occupied 840 plotting positions on the cathode ray tube (CRT) screen and subtended a visual angle of 2.1°. The horizontal and vertical position on the display scope of the centers of rotation of the objects was randomly varied by

±0.2°. The mean center-to-center separation of the objects was 3.8°. The eye-to-screen distance was 1.71 m. The dot and background brightnesses at the screen were approximately 5 cd/m^2 and 0.002 cd/m^2, respectively. A 0.5 neutral-density filter was inserted in the viewing tube to remove any apparent traces on the CRT.

Three models constructed from metal and plastic were used to instruct the subjects. Each model consisted of four white spheres connected by black rods. Two of the models were identical. The third model differed from the others only in the position of one of its spheres. The subjects responded by pressing one of two switches labeled *same* and *different*, respectively. The responses and response latencies were recorded by the PDP 11/44.

Procedure

Subjects were instructed to make *same* or *different* judgments for pairs of stimuli on the basis of the following criterion: "The two groups are the same when all of the distances between the dots are the same, regardless of their orientation. The two groups of dots are different when the distances between at least two of the dots are different." The three models were used to demonstrate the judgment criterion. Subjects who were to receive feedback were told that a single tone would indicate a correct response and that two successive tones would indicate an incorrect response. The room was darkened 2 min before the trials began.

Each subject participated in an initial screening session, 2 single-view sessions, 30 experimental sessions, 2 additional single-view sessions, and a final debriefing. Each session consisted of a baseline monitoring block and four experimental blocks. The baseline monitoring block consisted of two 12-view trials at each combination of levels of the points and motion variables, in random order. These trials were used to ensure that the subjects maintained a high level of accuracy when the number of views far exceeded the expected minimum levels. The 12-view trials were also intended to ensure that any failure to respond accurately at the minimum view levels was not due to the mental rotation component of the task, which should have been the same on the 12-view trials. Each experimental block consisted of 30 trials, each selected at random from the 60 possible conditions, so that there were 2 trials from each combination of levels of the three stimulus variables in each session.

There was a 1-min rest period between each block of trials. The order of the 34 sessions was randomized for each subject. Whichever sessions were selected as the first 2 and last 2 sessions for a given subject were run as single-view sessions by displaying only the first frame of each trial. (As with the dynamic displays, the frame was refreshed at 16 Hz until the subject responded.)

Results

The subject's task may be interpreted as that of determining whether there was a difference between the 3D structures represented by the standard and comparison stimuli. A signal detection paradigm (Green & Swets, 1966) was used to analyze the results, with the *different* trials serving as signal trials. A d' measure was computed for each subject and stimulus condition, using the proportion of "different" responses on *different* trials as the hit rate and the proportion of "different" responses on *same* trials as the false alarm rate. Each d' was based on 60 trials, half of which were signal (different) trials.

Single-View Trials

The independent variables in the analysis of the single-view trials were feedback, whether the single-view trials were presented before or after the dynamic trials, and number of points. There were two significant effects. The main effect of points, $F(3, 15) = 10.32$, $p < .01$, $\omega^2 = .151$, showed decreasing accuracy with increasing numbers of points. For the two-, three-, four-, and five-point conditions, the mean d's were 1.078, 0.478, 0.484, and 0.241, respectively. The interaction of feedback with the before versus after variable, $F(1, 5) = 7.51$, $p < .05$, $\omega^2 = .038$, showed a slight decrease in d' for the nonfeedback subjects (0.625 to 0.477) from the before trials to the after trials, but a larger increase (0.352 to 0.839) for the feedback subjects. This suggests that feedback, rather than mere exposure to the dynamic trials, improved performance on the single-view trials. The level of accuracy reached by the feedback subjects in the single-view sessions conducted after the dynamic sessions was comparable to that found in the two-view condition for those subjects.

The significance of the d' scores was calculated for each subject, for each combination of the stimulus variables, using Marascuilo's (1970, pp. 238–240) one-signal significance test. Of a total of 56 d's (7 subjects, before vs. after, and four numbers of points), 20 were significantly different from zero ($p < .05$). For the feedback subjects, 3 (of 12 total) d's were significant in the sessions prior to the dynamic trials, and 7 were significant in the later sessions. For the nonfeedback subjects, the number of significant d's was 5 (of 16 total) in both the early and late sessions. The greatest number of significant d's, 12 of 14 possible, occurred in the two-point condition.

There were no significant effects for the response bias, β, although there was a trend toward a larger "different" bias in the before trials. The mean βs for the before and after trials were 1.99 and 1.27, respectively.

An analysis of variance was conducted for the mean response latencies. (The latency analyses include trial type, same or different, as an additional independent variable. This variable does not appear in the d' and β analyses because both types of trials were used in computing those measures.) The only significant effect for latency was the main effect of points, $F(3, 15) = 12.15$, $p < .01$, $\omega^2 = .225$. The mean latencies for the two- through five-point conditions were 2.53 s, 4.20 s, 6.82 s, and 9.10 s.

The finding of above-chance performance in the absence of motion indicated that subjects were exploiting some regularity or regularities not included in the mathematical analyses of structure from motion. An analysis of the stimulus materials revealed the following relationships between characteristics of the simulated 3D objects and information in the 2D projections that could have resulted in above-chance accuracy in the single-view conditions: Pairs of two-point objects generated for *different* trials were necessarily different in 3D interpoint distance—this was the only definition possible for *different* objects in the two-point case. *Different* objects with more than two points also tended to differ in 3D interpoint distances as a result of the random displacement of a point used to generate differences in objects. This relationship diminished with increasing numbers of points. Although the controls of displaying the two objects at different initial slants and tilts and at different phases of rotation prevented a one-to-one correspondence between 3D interpoint distances and 2D interpoint distances, there was a correlation

between relative distances in 3D objects and relative distances in their 2D projections. This correlation, across our displays, was approximately .78. This resulted in a tendency for the interpoint distances in the projections of objects on *same* trials to be more similar than the interpoint distances in the projections of objects on *different* trials.[1] Although only 1 subject was aware of using this relationship, the availability of this regularity must be considered in interpreting the results of both the static and dynamic trials.

Dynamic Trials

The main effect of feedback was not significant for the dynamic trials, $F(1, 15) = 1.37, p > .05$, but there was a significant interaction of feedback with views, $F(4, 20) = 3.90, p < .05$, $\omega^2 = .007$. This interaction is illustrated in Figure 1. The feedback subjects show higher d's than the nonfeedback subjects for the smaller number of views, but this difference disappears at the six-view level. The main effect of views was significant, $F(4, 20) = 46.99, p < .01, \omega^2 = .116$. There was a significant main effect for points, $F(3, 15) = 23.63, p < .01, \omega^2 = .186$. Accuracy decreased with increasing numbers of points, with d's of 1.167, 0.944, 0.841, and 0.610, for the two- through five-point conditions, respectively.

The main effect of motion condition was significant, $F(2, 10) = 10.43, p < .01, \omega^2 = .037$, with d's of 0.992, 0.929, and 0.750, for the fixed-axis constant angular velocity, fixed-axis variable velocity, and variable-axis conditions, respectively. Post hoc comparisons (Tukey's honestly significant difference test) showed significant differences between the two fixed-axis conditions and the variable-axis condition, but not between the two fixed-axis conditions.

All interactions involving points, views, and motion conditions were significant. The F ratios and probabilities were as follows: $F(12, 60) = 3.63, p < .01$, for views with points; $F(8, 40) = 4.91, p < .01$, for views with motion; $F(6, 30) = 2.96, p < .05$, for points with motion; and $F(24, 120) = 2.09, p < .01$, for views with points with motion. The interaction of views with points is shown in Figure 2. The ω^2 values for these four interactions were .031, .040, .010, and .028, respectively. The interaction of points with motion condition, and of these two variables with views, is shown in Figure 3.

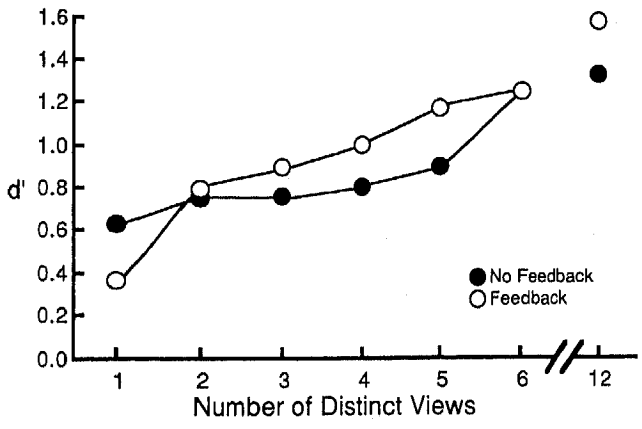

Figure 1. Effect of number of distinct views on d' for feedback and nonfeedback subjects. (The interaction described in the text refers to the two- through six-view conditions.)

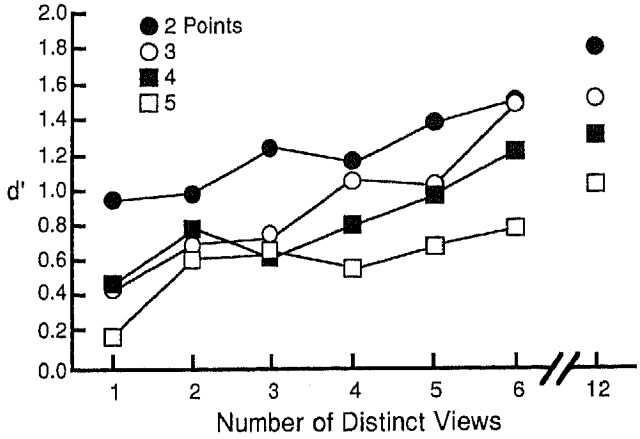

Figure 2. Interaction of number of points with number of distinct views. (The interaction described in the text refers to the two- through six-view conditions.)

The interaction of points with motion conditions shows a sharper drop in d' at the four- and five-point levels in the variable-axis condition, relative to the fixed-axis conditions. The interaction of motion condition with points and views is especially interesting to consider in detail because the structure-from-motion analyses are concerned with particular combinations of these three variables. Figure 3 shows that the sharper drop that occurs in d' after three points in the variable-axis condition, relative to the fixed-axis conditions, is especially apparent at five views. This is an indication of a possible critical combination of points, views, and motion conditions. Five views, according to Hoffman and Bennett's (1985) analysis, are sufficient for recovering 3D structure in two-point displays if the axis is fixed but not if the axis is variable. It is possible that the increased separation of the fixed-axis curves from the variable-axis curve at the five-view level is related to the availability of this additional information for recovering the depth coordinates of pairs of points (or for points in subgroups—see Discussion section), but this explanation must be regarded as speculative until confirmed by additional research. Separate inspection of the proportions of correct responses for the *same* and *different* trials indicates that the interaction of points with motion constraints can be attributed to the *same* trials. This is illustrated in Figure 4 for the five-view condition.

The numbers of d' values for individual subjects that were significantly different from zero ($p < .05$) in the two- through

[1] The means (and standard deviations) of the differences between the two objects in a pair in 2D RMS interpoint distances (in radius units) for same and different trials, respectively, were 0.35 (0.32) and 0.72 (0.39) for two-point objects; 0.25 (0.20) and 0.46 (0.29) for three-point objects; 0.20 (0.15) and 0.25 (0.18) for four-point objects; and 0.18 (0.14) and 0.21 (0.15) for five-point objects.

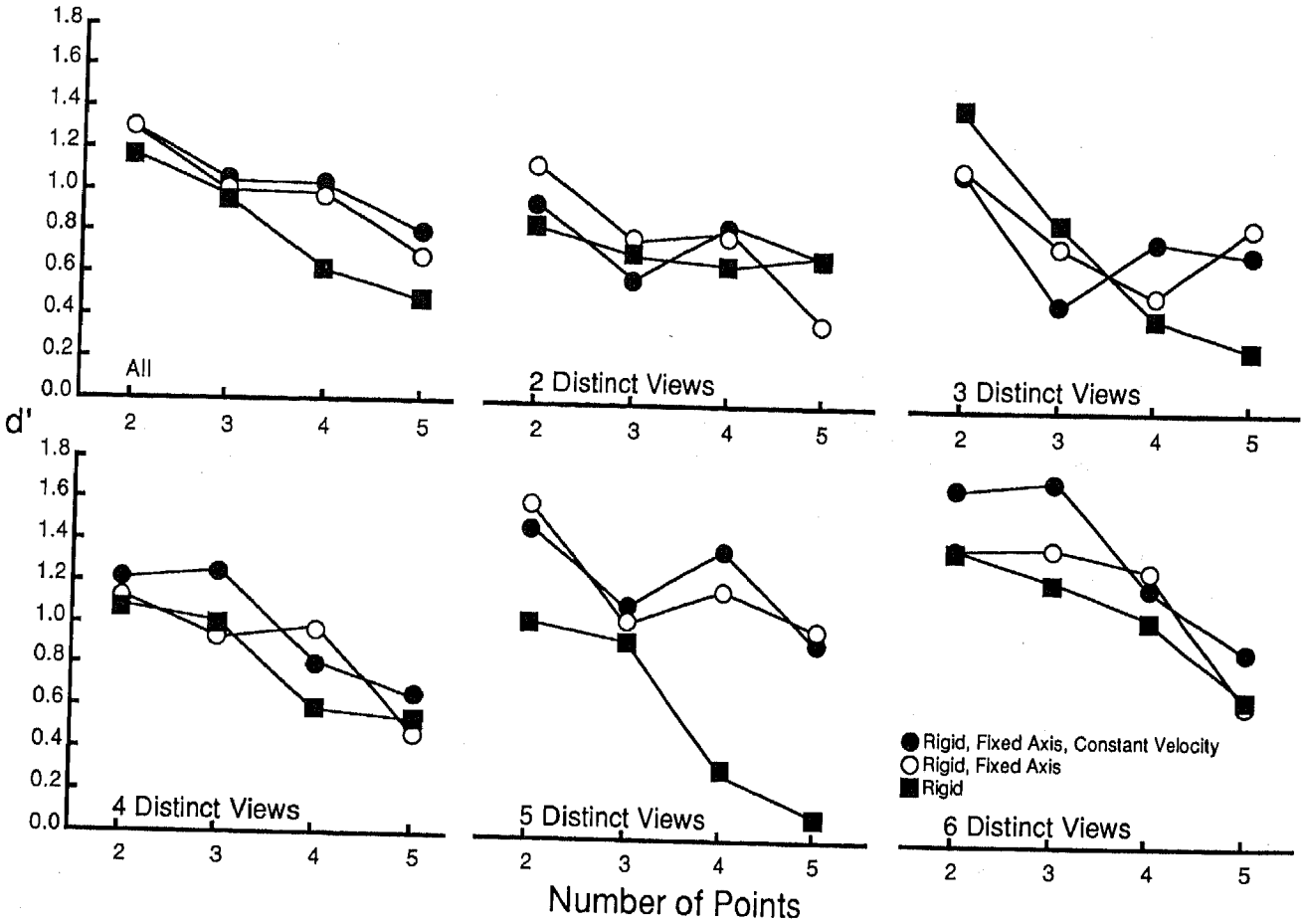

Figure 3. Interaction of number of points with motion condition for two through six views.

six-view conditions for the feedback and nonfeedback subjects, respectively, were 138 (of a total of 180 possible) and 154 (of a total of 240 possible). There was no indication of changes from below-chance to above-chance performance at any of the critical levels in the structure-from-motion analyses.

The mean of the response bias, β, for the two- through six-view conditions was 1.050. There were no significant effects of the independent variables on β.

An analysis of the response latencies for the two- through six-view conditions showed an increase in latency with increasing numbers of points, $F(3, 15) = 13.93$, $p < .01$, $\omega^2 = .311$. The mean latencies were 2.48 s, 3.98 s, 6.74 s, and 9.15 s for the two-, three-, four-, and five-point conditions. The increase between three and five points was nearly linear. A more rapid increase with number of points would be expected if subjects were examining each interpoint distance sequentially. There was a significant interaction of points, views, and motion conditions, $F(24, 120) = 1.70$, $p < .05$, $\omega^2 = .001$. There were no other significant effects. If subjects compared all interpoint distances across displays on *same* trials, but responded as soon as one nonmatching distance was found on *different* trials, there should have been an interaction of number of points with type of trial. This interaction was not significant, $F(3, 15) = .18$. A comparison of this value with the large main effect found for number of points indicates that this type of differential processing on *same* and *different* trials did not occur.

Separate analyses were conducted for d', β, and latencies for the 12-view trials. For the d' values, the main effect of points was significant, $F(3, 15) = 10.78$, $p < .01$, $\omega^2 = .171$. The mean d's for the two- through five-point displays were 1.854, 1.568, 1.348, and 1.085, respectively. The main effect of motion type was significant, $F(2, 10) = 18.42$, $p < .01$, $\omega^2 = .314$. The mean d's were 1.641, 1.810, and 0.941, for the fixed-axis with constant velocity, fixed-axis, and variable-axis conditions, respectively. The fixed-axis conditions were significantly different from the variable-axis condition ($p < .01$) but not from each other. The main effect of feedback was not significant, $F(1, 5) = 3.55$, $p > .05$. There were no significant interactions.

The β analysis for the 12-view trials revealed a significant effect of the motion condition, $F(2, 10) = 7.34$, $p < .05$, $\omega^2 = .048$, and a significant interaction of motion with points, $F(6, 30) = 2.81$, $p < .05$, $\omega^2 = .024$. The mean β values were .77 for both of the fixed-axis conditions, indicating a *same* bias, and 1.01 for the variable-axis condition. The bias occurred for the smaller numbers of points, disappearing at the five-point level.

Latencies in the 12-view trials increased with number of

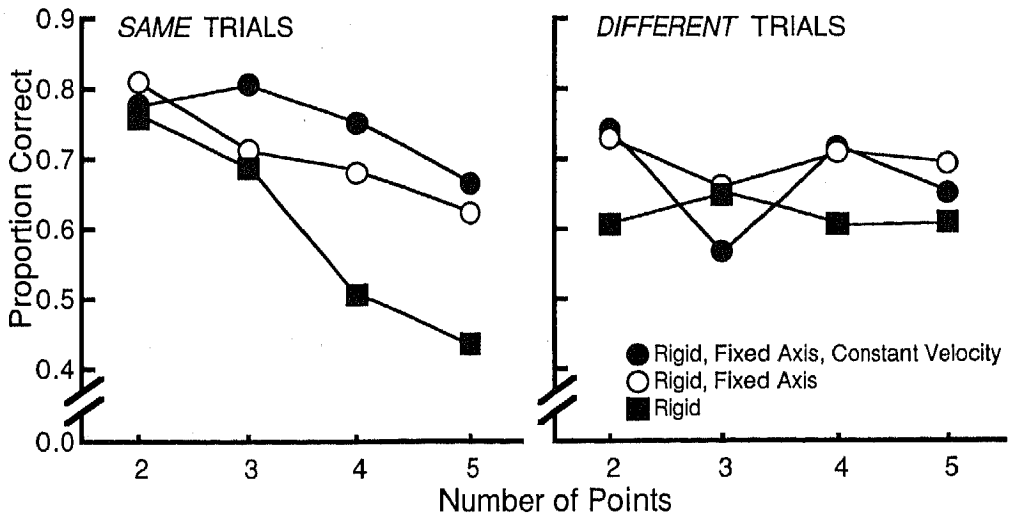

Figure 4. Interaction of number of points with motion condition at five views, for *same* and *different* trials.

points, $F(3, 15) = 13.09$, $p < .01$, $\omega^2 = .222$. Mean latencies for two through five points were 2.58 s, 3.52 s, 5.44 s, and 6.99 s. Latencies were also affected by the motion condition, $F(2, 10) = 10.61$, $p < .01$, $\omega^2 = .002$. The mean latencies for the fixed-axis conditions, 4.61 and 4.56 s, were significantly different from the 4.84-s mean latency for the variable-axis condition.

In response to debriefing questions, 5 of the 7 subjects reported seeing the simulated objects as 3D on 100% of the trials in both the static and dynamic conditions. One subject reported seeing the objects as 3D on 50% of the static trials and on 90% of the dynamic trials. The remaining subject reported seeing the objects as 3D on 30% of the static trials and 85% of the dynamic trials. In response to a question concerning strategies used in the experiment, 1 subject (the one who was not naive) reported using the differences in the projected interpoint distances in the two objects to make judgments for the two-point displays. The other subjects reported mental rotation in 3D as the only strategy of which they were aware.

Discussion

Theoretical analyses of the recovery of 3D structure from 2D images have shown that three views of four points are sufficient to recover structure under a rigidity constraint (Ullman, 1979), with fewer points required as additional constraints and/or views are added (Hoffman & Bennett, 1985, 1986). As these proofs assume infinitely fine resolution and the absence of noise, one might expect poorer performance by human observers. More views should be required if an incremental rigidity scheme (Ullman, 1984) is used to overcome the effects of noise. Human performance in our experiments might be expected to be degraded further because of the interposition of a task possibly requiring mental rotation between the recovery of structure and the behavioral response. For these reasons, our finding that subjects could make accurate psychophysical judgments with fewer points and distinct views than expected on the basis of theoretical analyses was especially surprising. It should of course be emphasized that the theoretical analyses are concerned with recovery of the 3D coordinates of points in an arbitrarily scaled space, up to a reflection about the image plane, whereas our subjects were comparing pairs of structures. The implications of these differences between the theoretical concept of recovery of structure and the requirements of our behavioral task are discussed below.

It is clear that accurate responses in our comparison task did not depend entirely on motion. Subjects performed above chance levels when presented with static views. This result indicates that subjects exploited regularities that are not included in the structure-from-motion analyses considered in this article. The principal regularity is likely to have been the correlation between the 3D and 2D interpoint distances that occurs across objects that vary randomly in the positioning of points in 3D. This relationship, which was greatest for two-point objects and decreased with increasing numbers of points, is likely to have contributed to the decrease in accuracy that occurred with increasing numbers of points. Another factor that may have contributed to that decrease is the increase in complexity of the structure, in terms of the number of interpoint distances, that occurs with greater numbers of points. Any task used to ascertain whether the relative depth coordinates of points have been correctly recovered is likely to be more difficult for greater numbers of points.[2] This appears to have been the case for the comparison task in the present experiment. This conclusion is supported by the latency results: Response time increased as the number of points in the structures increased.

[2] The theoretical analyses considered in the present study were concerned with recovering depth coordinates for individual points. The task that we used would not be appropriate for studying analyses concerned with recovering surface structure (e.g., Koenderink & van Doorn, 1986). Different results might be expected for number of points if the task involved detection of surfaces or discrimination among surfaces (Uttal, 1987).

Subjective reports of the appearance of the stimuli suggest several other regularities that might have been exploited by the subjects. In the theoretical analyses considered in this article, a point in a static display (of the type used in the present study) could represent any location along a line extending in depth for an infinite distance. A human subject, on the other hand, may perceive a point as located at a specific distance and may perceive adjacent points as being equally distant (Gogel, 1973). Subjects may also tend to perceive the extent in depth of an object as being of approximately the same magnitude as that of the perceived height and width of an object. This would be a reasonable heuristic in a natural environment and would have been appropriate to our displays, which were based on spherical objects, and to most other displays studied in structure-from-motion research.

Subjects also may have exploited the constraint of a constant scale in relating distances in the projection to the 3D distances that were represented. Scale is undetermined in the mathematical analyses, and the subjects indeed may impose an arbitrary scale in recovering the 3D structures of the displays. It seems unlikely that different scales would be imposed on different displays, however, especially for displays within the same pair. The assumption that the scale is the same for both displays in a pair is essential to accurate responding to the two-point displays. Indeed, we could not have scored a *same* or *different* response to a pair of two-point displays as correct or incorrect without assuming equal scales.

We found a greater decrease in accuracy with increasing numbers of points with variable-axis motion than with fixed-axis motion. This interaction was primarily due to the *same* trials. A possible reason for this is the difference in the requirements for a correct response on *same* trials as compared with *different* trials. To verify that two structures are the same, each structure must be recovered uniquely. To determine that two structures are different, it is only necessary to identify a set of possible structures for each display and to determine that these two sets do not overlap.

The drop in accuracy with increasing numbers of points on *same* trials was especially marked after three points for the variable-axis condition. Verbal reports indicated that the subjects attempted to organize the display into subunits of no more than three points. A four-point display might be perceived as a triangle and a dot, a five-point display as a triangle and a rod. Our hypothesis is that it was more difficult to maintain a perception of rigid relationships among subunits for the variable-axis displays. The use of triangular subunits by subjects in these judgments, and the importance of triangles in the analysis of optic flow (Koenderink & van Doorn, 1986), may be more than coincidental.

These suggestions of possible grouping effects indicate that organization of feature points into subgroups should be examined as a potentially important component of the recovery of 3D structure from dynamic 2D images. Some principles for grouping based on orthographic projections of rotation in depth have been reported by Gillam (1976). In studies of the recovery of structure from motion, it would be important to determine whether grouping was based on 3D or on 2D relationships. This might indicate whether grouping or recovery of structure occurs first or perhaps would show that the two processes occur in parallel.

The higher level of accuracy found with fixed-axis rotation is consistent with the theoretical analyses (Hoffman & Bennett, 1985) which show that fewer points and views are required with fixed-axis motion than with variable-axis motion. Our results parallel Todd's (1982) finding of greater accuracy in discriminating rigid from nonrigid motion with fixed-axis than with variable-axis motion. The present finding of increasing accuracy with increasing numbers of distinct views is consistent with Todd's (1982) trajectory-based analysis. We did not distinguish between distinct views and length of the trajectory. A greater number of distinct views displayed a greater extent of the motion trajectory. Although one could use different numbers of 2D frames to display a given extent of a trajectory in an artificial display, it is not clear that separating the concepts of distinct views and extent of the motion trajectory is useful in studying the perception of continuous motion by human observers. The issue of distinct views would be an interesting topic for further study.

In conclusion, the present study used an indirect method to test theories about the recovery of structure from motion that are not directly testable. The theories considered in this article are theories about competence (Ullman, 1986), and it may be necessary to elaborate these theories to include direct implications for human performance if they are to be subjected to direct psychophysical testing. It may also be necessary to develop psychophysical techniques for the study of dynamic information for depth perception to supplement current techniques, which emphasize detection of minimal differences. It is often of importance to determine how different stimuli are classified by an observer (e.g., as 2D vs. 3D or as rigid vs. nonrigid), even when the stimuli are discriminable on other dimensions. As such developments in mathematical analysis and in psychophysics proceed, it should become possible to combine mathematical and psychophysical approaches to the study of more complex patterns of optic flow (Koenderink, 1986).

References

Bennett, B., & Hoffman, D. (1985). The computation of structure from fixed axis motion: Nonrigid structures. *Biological Cybernetics, 51,* 293–300.

Braddick, O. (1974). A short-range process in apparent motion. *Vision Research, 14,* 519–527.

Braunstein, M. L., & Andersen, G. J. (1986). Testing the rigidity assumption: A reply to Ullman. *Perception, 15,* 641–644.

Gillam, B. (1976). Grouping of multiple ambiguous contours: Towards an understanding of surface perception. *Perception, 5,* 203–209.

Gogel, W. C. (1973). The organization of perceived space: I. Perceptual interactions. *Psychologische Forschung, 36,* 195–221.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics.* New York: Wiley.

Hoffman, D., & Bennett, B. (1985). Inferring the relative 3-D positions of two moving points. *Journal of the Optical Society of America, 75,* 350–353.

Hoffman, D., & Bennett, B. (1986). The computation of structure from fixed axis motion: Rigid structures. *Biological Cybernetics, 54,* 1–13.

Hoffman, D., & Flinchbaugh, B. (1982). The interpretation of biological motion. *Biological Cybernetics, 42,* 197–204.

Koenderink, J. J. (1986). Optic flow. *Vision Research, 26,* 161–180.

Koenderink, J. J., & van Doorn, A. J. (1986). Depth and shape from differential perspective in the presence of bending deformations. *Journal of the Optical Society of America, 3,* 242–249.

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.

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

Marascuilo, L. A. (1970). Extensions of the significance test for one-parameter signal detection hypotheses. *Psychometrika, 35,* 237–243.

Schwartz, B. J., & Sperling, G. (1983). Nonrigid 3D percepts from 2D representations of rigid objects. *Investigative Ophthalmology and Visual Science, 24*(3, Supplement), 239.

Stevens, K. A. (1983). Surface tilt (the direction of slant): A neglected psychophysical variable. *Perception & Psychophysics, 33,* 241–250.

Todd, J. (1982). Visual information about rigid and nonrigid motion: A geometric analysis. *Journal of Experimental Psychology: Human Perception and Performance, 8,* 238–252.

Todd, J. (1985). The perception of structure from motion: Is projective correspondence of moving elements a necessary condition? *Journal of Experimental Psychology: Human Perception and Performance, 11,* 689–710.

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.

Ullman, S. (1986). Competence, performance, and the rigidity assumption. *Perception, 15,* 644–646.

Uttal, W. R. (1987). *The perception of dotted forms.* Hillsdale, NJ: Erlbaum.

Webb, J. A., & Aggarwal, J. K. (1981). Visually interpreting the motion of objects in space. *Computer, 14*(8), 40–46.

Received December 31, 1986
Revision received January 20, 1987
Accepted January 20, 1987 ■

Correction to Connine and Clifton

In the article "Interactive Use of Lexical Information in Speech Perception" by Cynthia M. Connine and Charles Clifton, Jr. (*Journal of Experimental Psychology: Human Perception and Performance,* 1987, Vol. 13, pp. 291–299), Figures 1 and 2 were inadvertently transposed. The figure on p. 294 is actually Figure 2, and the figure on p. 296 is actually Figure 1. The captions are correct as they stand.
