Perception, 1993, volume 22, pages 1441-1465

Inferring structure from motion in two-view and multiview displays

Jeffrey C Liter, Myron L Braunstein, Donald D Hoffman
Department of Cognitive Sciences, University of California, Irvine, CA 92717, USA
Received 23 March 1993, in revised form 17 September 1993

Abstract. Five experiments were conducted to examine constraints used to interpret structure-from-motion displays. Theoretically, two orthographic views of four or more points in rigid motion yield a one-parameter family of rigid three-dimensional (3-D) interpretations. Additional views yield a unique rigid interpretation. Subjects viewed two-view and thirty-view displays of five-point objects in apparent motion. The subjects selected the best 3-D interpretation from a set of 89 compatible alternatives (experiments 1-3) or judged depth directly (experiment 4). In both cases the judged depth increased when relative image motion increased, even when the increased motion was due to increased simulation rotation. Subjects also judged rotation to be greater when either simulated depth or simulated rotation increased (experiment 4). The results are consistent with a heuristic analysis in which perceived depth is determined by relative motion.

1 Introduction

A fundamental problem long recognized in vision research is the one-to-many inverse relation between retinal images and the visual environment. In general, for a particular retinal projection, there are an infinite number of potential three-dimensional (3-D) interpretations of the environment which are consistent with that projection. The different interpretations are related by manipulations of the depth along lines of sight. In order to deal with the inherent ambiguity of retinal projections, human vision uses geometric and environmental regularities. Exploiting some of these regularities may require integration of image information over time. For example, humans can recover aspects of the 3-D shape of an object from projected motion of features on the surface of the object. This ability, termed the 'kinetic depth effect' by Wallach and O'Connell (1953) and 'structure from motion' (SFM) by Ullman (1979), has been the subject of many empirical and theoretical studies (for a review of the earlier empirical literature, see Braunstein 1976).

Visual motion does not by itself solve the projective ambiguity (see, for example, Johansson 1970). Some assumptions or constraints are also necessary. The minimum number of feature points and retinal images required to recover 3-D structure from motion has been determined under constraints of rigidity (Ullman 1979), pairwise rigidity and planarity (Hoffman and Flinchbaugh 1982), rigidity and fixed-axis motion (Hoffman and Bennett 1985), and fixed-axis motion only (Bennett and Hoffman 1985). These analyses, which we refer to as computational analyses, have two distinguishing properties. (1) If their constraints are satisfied by the objects being viewed then they recover the correct 3-D structure of the objects (up to a reflection about a plane perpendicular to the line of sight for orthographic projections). (2) The probability of false targets is zero. (A false target would occur, for example, if an analysis requiring rigidity found a rigid solution to a nonrigid motion.) Other analyses have been proposed which do not guarantee a correct solution and do not have zero probability of false targets. We refer to these analyses as heuristic analyses (see, for example, Braunstein 1976, 1993). Constraints that fail to eliminate false targets or fail to provide veridical solutions (when the constraints are satisfied) may be inappropriate for a computational analysis, but may still be used in a heuristic analysis.

Recent theoretical investigations have found that two distinct orthographic views of a rigid object yield a one-parameter family of rigid 3-D interpretations of the object's structure[1] (Aloimonos and Brown 1989; Bennett et al 1989; Huang and Lee 1989; Koenderink and van Doorn 1991; Todd and Bressan 1990). Thus, even under ideal conditions, a computational analysis in which just a rigidity constraint is used cannot recover the original 3-D structure of the object. In fact, in order to recover a unique structure one must use additional constraints. Recent research with two-view orthographic displays (Braunstein et al 1987, 1990) suggests that subjects perceive some 3-D structure in these displays. Similarly, the experiments of Todd and his colleagues suggest that subjects perceive some 3-D structure (not necessarily metric in nature) in two-view SFM displays (Todd and Bressan 1990; Todd and Norman 1991; Todd et al 1988). These experiments, however, do not address the question of how the structure perceived in these displays is systematically related to characteristics of the two-dimensional (2-D) images. In the experiments which follow, we address this question. We then examine some heuristic processes which may explain our results.

The remainder of the paper is organized as follows. We first discuss the theoretical findings concerning the recovery of structure from two orthographic views. We then present five psychophysical experiments. In the first three experiments subjects observed a two-view or a multiview display of a simulated 3-D object and selected a rigid 3-D interpretation from the one-parameter family of such interpretations described by Bennett et al (1989). In experiment 1 we considered whether specific rigid 3-D interpretations are selected for two-view displays. In experiment 2 we considered multiview displays, for which a unique interpretation (plus reflection) is theoretically recoverable. The results of experiments 1 and 2 indicated that subjects used constraints in addition to rigidity to interpret 3-D structure both in two-view displays and in multiview displays, and that these constraints might involve assumptions about the depth or rotation magnitude of the simulated objects. In experiment 3 we varied the simulated depth and the simulated rotation angle in the objects used to generate the displays. We found that interpretations with greater depth were selected as the relative motion among the image features increased, whether this increase was due to increased simulated depth or increased simulated rotation magnitude. (We provide a precise definition of relative motion in section 5.2.) In the remaining experiments we obtained direct judgments of perceived depth (experiment 4) and perceived rotation (experiment 5) in two-view and multiview oscillating displays and examined the relationship of these judgments to relative motion.

2 Two-view structure from motion

Recent computational analyses of the two-view SFM problem conclude that if there is any rigid 3-D interpretation of a two-view SFM display, then there is, in fact, an infinite family of rigid 3-D interpretations[2] (Aloimonos and Brown 1989; Bennett et al 1989; Huang and Lee 1989; Koenderink and van Doorn 1991; Todd and

[1] Two-view perspective projections provide sufficient information for recovering 3-D structure, and accurate responses to such displays have been reported by Doner et al (1984) and Lappin et al (1980). For brevity we use SFM in this paper to refer exclusively to orthographic projections of rotations.

[2] The members of the family are related by a single parameter. If the two views are sufficiently close in space and time to approximate a velocity field, then the members of the family are related by an affine stretching along the line of sight (Hoffman 1982; Todd and Bressan 1990; Ullman 1983). For two discrete views, the relationship among the members of the family is more complex (Aloimonos and Brown 1989; Bennett et al 1989; Huang and Lee 1989; Koenderink and van Doorn 1991).

Bressan 1990; see also Hoffman 1982; Ullman 1983). We focus our discussion on the family of interpretations described by Bennett et al (1989).

The goal of their analysis was to recover the 3-D coordinates $(x_{i,j}, y_{i,j}, z_{i,j})$ of a collection of n image feature points viewed in m time frames (here, $i = 1, ..., n$ indexes image features and $j = 1, ..., m$ indexes time frames). They assume that the projection is orthographic and that the correspondence among image feature points is known from one time frame to the next. The $x_{i,j}$ and $y_{i,j}$ are trivially available from the image, but to determine the $z_{i,j}$, they integrate information over multiple time frames and use the assumption of rigid motion. Bennett et al (1989) prove that, with one point foveated to eliminate translation parallel to the image plane and with the assumption of a rigid 3-D rotation, two views of four image feature points determine a one-parameter family of depths $z_{i,j}$. Considering more than four feature points does not resolve the ambiguity. Call the free parameter t. For each t there is a rigid 3-D interpretation, and the 3-D transformation Ψ_t relating frames 1 and 2 of this interpretation can be determined. Ψ_t is a 3-D rotation and one can determine the rotation axis and angle for each t. The tilt of the rotation axis (ie the angle between the projection of the axis into the image plane and the X-axis) is the same for all t and the slant of the rotation axis varies from 0° (rotation axis in the image plane) to 90° (rotation axis along the Z-axis, ie along the line of sight) as t varies from 0 to ∞. For $t = 0$ the z_{ij} are undefined so that 0° slant is not attained. The rotation angle approaches its maximum (180°) as t approaches 0 and approaches its minimum as t approaches ∞. For the remainder of the paper, we will use the slant of the rotation axis to index the various 3-D interpretations of a two-view display. This is equivalent to specifying t but has a more intuitive geometric interpretation.

For each display in the experiments that follow, we began with a 3-D object composed of five points in 3-space (the 'generating object'). We chose two planes of projection and from these two projections obtained a one-parameter family of 3-D interpretations. For each interpretation in this one-parameter family we can compute a 3-D rotation angle (the angle of rotation between the two views in that interpretation) and a line-of-sight depth. We compute the line-of-sight depth for a particular 3-D interpretation in the one-parameter family as follows: (1) orient the 3-D interpretation so that it projects to frame 1 of the two-view display; (2) compute the maximum difference of the z (depth) coordinates; (3) orient the 3-D interpretation so that it projects to frame 2 of the two-view display; (4) compute the maximum difference of the z coordinates; (5) average the values obtained in steps (2) and (4). We will use the line-of-sight depth, computed in this way, to characterize the depth in the 3-D interpretations selected by subjects in experiments 1–3. Figure 1 shows how the

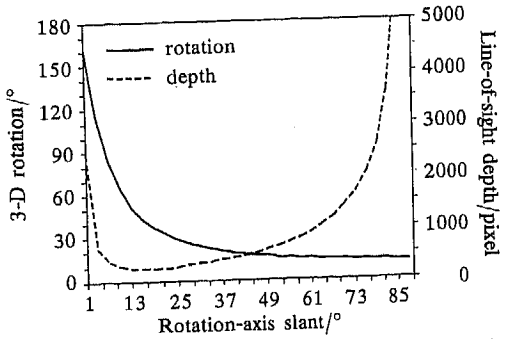

Figure 1. The line-of-sight depth and 3-D rotation between end views for the 3-D interpretations of display 6 studied in experiments 1 and 2.

line-of-sight depth and the 3-D rotation angle vary with rotation-axis slant (and thus with t) for the 3-D interpretations obtained from one of the two-view displays used in experiments 1 and 2.

3 Experiment 1

As discussed in the previous section, if there is any rigid 3-D interpretation of a two-view orthographic SFM display, there are an infinite number of rigid 3-D interpretations. Still, some 3-D structure is perceived in two-view orthographic displays. In the present experiment, subjects viewed an oscillating two-view display and selected that 3-D interpretation from the family of mathematically compatible interpretations described by Bennett et al (1989) which they believed had the 3-D structure most similar to that perceived in the two-view display. Our objective was to find any regularities to the responses that might indicate what constraints subjects were using to disambiguate these displays.

3.1 *Method*

3.1.1 *Subjects*. The subjects were three graduate students from the University of California, Irvine, who were paid for their participation. All of the subjects were naive to the purposes in the experiment and had normal or corrected-to-normal vision (20/40, Snellen eye chart).

3.1.2 *Design*. The slant of the rotation axis which was used to generate the two-view display (the 'generating slant') was the only independent variable. One display was produced at each of sixteen slant angles (see section 3.1.3).

3.1.3 *Stimuli*. The stimuli were white dots moving on a dark background, simulating orthographic projection of an object undergoing rigid fixed-axis rotation. In the 3-D model used to produce the display, five points were randomly selected within the volume of a sphere of radius 512 pixels[3] (6.3 deg of visual angle), subject to the constraints discussed below. The first frame of the motion sequence was an orthographic projection of these points. A 3-D rotation axis and rotation angle were randomly chosen to produce the second frame of the motion sequence. The tilt of the rotation axis was selected randomly between 0° (parallel to the X-axis) and 180°. The slant of the rotation axis was chosen randomly without replacement from a set of sixteen angles. The set contained two randomly selected angles from each of eight 11° intervals starting from 1° (nearly in the image plane) and increasing to 88° (nearly along the line of sight). Specifically, two angles were selected at random from the interval [1, 11], two from [12, 22] and so on. The rotation angle, in degrees, between the two views was selected at random from the interval [8, 18] for each display. All of the subjects were presented with the same 16 displays.

Several restrictions were placed on the 3-D locations of the five points in a generating object. A nearest-neighbors constraint was used to prevent correspondence mismatches. For every dot in the first frame of the motion sequence, the projected 2-D distance between that dot and every noncorresponding dot in the second frame of the motion sequence was required to be at least 5% greater than the distance between that dot and its corresponding dot in the second frame. To prevent stationary or nearly stationary dots in a display, each dot in the projection was required to move by at least 7.5% of the maximum radius (38 pixels) in the transition from frame 1 to frame 2. To assure that the dots were spatially distributed in the

[3] The term pixel does not strictly apply to the calligraphic displays used in these experiments but is used as a convenient shorthand for 'plotting position', the number of distinct locations at which the center of a dot can be positioned. The same pixel units are used in the Z (depth) dimension to describe the geometric model used to generate the projections.

displays, dots in the 2-D projection were not allowed to be within 20% of the maximum radius (102 pixels) of one another and points in the 3-D model were not allowed to be within 50% of the maximum radius (256 pixels) of one another. On average ten points were eliminated while generating each display in order to meet these criteria. 16 displays were chosen in this way, 1 for each of the slant values.

For each of the 16 two-view displays, we created a corresponding set of 89 'response displays'. Each subject was to select from this set of 89 response displays the display depicting a 3-D structure most similar to that of the two-view display. Each response display consisted of 450 frames, depicting a full 360° of rigid rotation about a fixed axis. These 450 frames were repeated in order (1, 2, ..., 450, 1, 2, ...) for as long as a subject wished to view them, giving an impression of continuous rigid rotation about a fixed axis. Each response display depicted a rigid 3-D structure from the one-parameter family of such structures compatible with the two-view display (see Bennett et al 1989). The first response display depicted the rigid structure whose axis of rotation had a slant of 1°; the second depicted the rigid structure whose axis of rotation had a slant of 2°; the same pattern continued through the 89th response display, which depicted the rigid structure whose axis of rotation had a slant of 89°. In this fashion, the set of 89 response displays sampled the full range of rigid 3-D interpretations contained in the one-parameter family corresponding to the two-view display.

On each trial, one of the 16 two-view displays was situated on one side of the display oscilloscope (left or right, counterbalanced across subjects). It oscillated back and forth between the two views with an SOA of 400 ms (see Todd et al 1988). Another display, the response display, was situated on the other side of the display oscilloscope. As described above, it depicted continuous fixed-axis rotation of one of the 3-D interpretations consistent with the two-view display. The 3-D rotation rate of the response display was 0.8° per frame. The refresh rate and update rate were 75 frames s^{-1}. The centers of the two displayed objects were separated by 12.6 deg of visual angle.

3.1.4 *Apparatus*. The displays were presented on an IMI vector graphics system with 4096 × 4096 pixel resolution. Subjects viewed the displays monocularly from a distance of 100 cm through a reduction tube and round aperture which limited the field of view to a circular region 3072 pixels in diameter (a visual angle of 18.7 deg). The reduction tube was fitted with a 0.4 neutral-density filter which helped to eliminate visible traces on the CRT and produced a darker background. The subjects used a force joystick to move through the set of 89 response displays in search of the best match. Unless otherwise noted, the apparatus was identical for the remaining experiments.

3.1.5 *Procedure*. The subjects were run individually in two sessions. The first session consisted of twenty-four trials—eight practice trials all different from the experimental trials followed by a random ordering of the sixteen experimental trials. In the second session there were twenty trials—four practice trials followed by a different random ordering of the sixteen experimental trials.

The subjects were instructed that their task was to "match two three-dimensional objects", by using the joystick "to change the structure of the continuously rotating object so as to match its structure to that" of the oscillating object. They were to press the button on the joystick when they were satisfied that they had found "the best match". At the beginning of each trial, one of the 89 response displays, depicting a 3-D interpretation consistent with the two-view display, was selected at random and presented next to the two-view display. When the subject pushed the joystick forward, a new response display, indexed by a larger rotation-axis slant, replaced the old response display. When the joystick was pulled backward, a new response display indexed by a

smaller rotation-axis slant replaced the old response display.[4] Except when the subject was manipulating the joystick, the projected motion of the response display depicted fixed-axis rotation of a rigid 3-D object.

The subjects were allowed to search through the 89 response displays for as long as was necessary and were encouraged to be as precise as possible. The subjects were informed that the response display might appear nonrigid while they were making adjustments with the joystick and that they should watch it rotate without making adjustments when comparing it with the oscillating display. To assure that subjects viewed all possible response displays before responding, they were required to move through the full range of response displays on each trial by adjusting the joystick until the limit was reached in both directions, 1° of slant and 89° of slant. When either limit was reached, the computer terminal in the experimental room beeped. Because of the extreme relative depth depicted in some of the response displays near the limits of the slant range, some of the dots in some of the response displays did not remain on the display oscilloscope for the entire 360° rotation cycle. When we present the results, we will identify the range of slants for which all dots remained on the display oscilloscope for the entire 360° rotation cycle.

3.2 *Results and discussion*

As noted above, the slant of the rotation axis can be used to index the family of rigid 3-D interpretations of a two-view display. To examine regularities in the 3-D interpretations that subjects selected for the two-view displays, we compared their selected interpretations for each display with the 3-D interpretations that we used to generate the display. Specifically, for each of the 16 displays, we compared the average slant index of the two judgments with the generating slant. All three subjects showed a significant positive correlation ($p < 0.05$), $r_{14} = 0.799$, 0.725, and 0.918 for subjects 1, 2, and 3, respectively. The relationship between the selected interpretation and the generating interpretation is shown in figure 2 for each subject. The shaded regions indicate the 3-D interpretations for which not all of the feature points remained on the display oscilloscope for the entire rotation cycle.

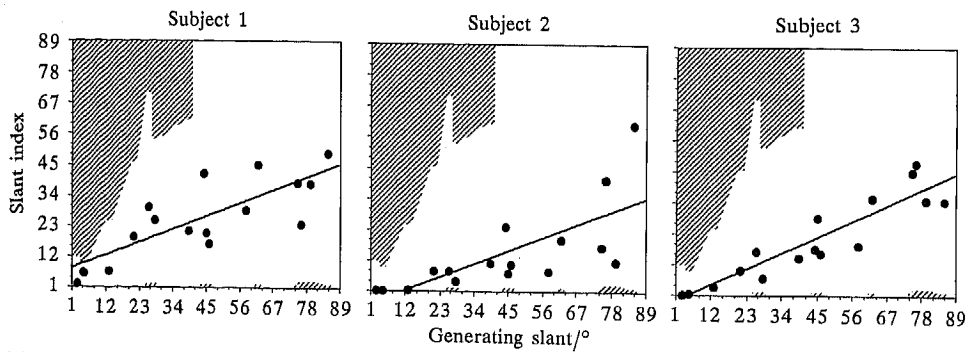

Figure 2. The correspondence between slant index of the selected interpretation and generating slant in experiment 1. The shaded regions denote interpretations (response displays) for which not all of the dots remained on the display oscilloscope for the complete rotation cycle.

[4] To make the transition between response displays as smooth as possible when the subject manipulated the joystick, the sequencing of views was always maintained. For example, frame number 140 (of the total 450 frames) was always followed by frame number 141, regardless of whether a new response display was selected in the time interval between the two frames. Frame number 141 always depicted the 3-D interpretation as it would appear rotated 112° from the orientation which projected to frame 1 of the two-view display. Since the 3-D structure is similar for interpretations with similar rotation-axis slants (see figure 1), the transition between response displays usually appeared to be a deformation of a single rotating object rather than replacement by a new object.

The relationship between the selected interpretation and the generating interpretation shown in figure 2 is surprising since the information in a two-view orthographic projection is insufficient mathematically to recover the generating interpretation. With rigidity the only constraint (see, for example, Bennett et al 1989), all of the 89 interpretations that were displayed on each trial were equally consistent with the two-view display. The results of this experiment suggest that human observers use constraints in addition to rigidity to select a particular interpretation. The correlation between the slant index of the selected interpretation and the slant of the generating interpretation further suggests that these constraints may be similar to or related to those that were used to generate the two-view displays. For example, the two-view displays were generated in a spherical volume and with small rotation angles (8° to 18°). A bias to see objects as no deeper than they are wide [ie a compactness constraint (Proffitt et al 1992)] or to see objects which rotate through small angles might produce a correlation between the selected interpretations and the generating interpretations. We examined the basis for this relationship in experiments 3, 4, and 5 by systematically varying the depth and rotation angle in the generating objects.

Although slant provides a precise indexing of the 3-D interpretations, examining the depth simulated in the selected interpretation as a function of the generating slant may also be interesting. We plotted the line-of-sight depth (see section 2) in each of the selected interpretations as a function of the generating slant. All three subjects showed a significant correlation ($r_{14} = -0.812$, -0.563, and -0.746, respectively, $p < 0.05$). The depth in the selected interpretations decreased as the slant of the rotation axis used to generate the two-view displays approached the viewing direction. This relationship is displayed in figure 3 for all three subjects. The depth simulated in the selected interpretations was, on average, less than the depth simulated in the generating interpretations. The difference between the mean simulated depth and mean depth in the selected interpretations was 192 pixels for subject 1, 273 pixels for subject 2, and 301 pixels for subject 3.

Because of our random generation procedure, there was no systematic relationship between generating slant and the maximum projected width of the two-view displays ($r_{14} = 0.111$, $p > 0.05$) but there were variations in projected width across the different displays. The average display width was 795 pixels and the standard deviation over the 16 displays was 78 pixels. Contrary to what might be expected on the basis of a compactness constraint, there was no relationship between the depth in the selected interpretations and the projected width of the two-view displays ($r_{14} = 0.146$, 0.028, and 0.095 for subjects 1, 2, and 3, respectively). This result is not necessarily inconsistent with a compactness constraint, however, since it is possible that the width variations in the present experiment were too small to affect

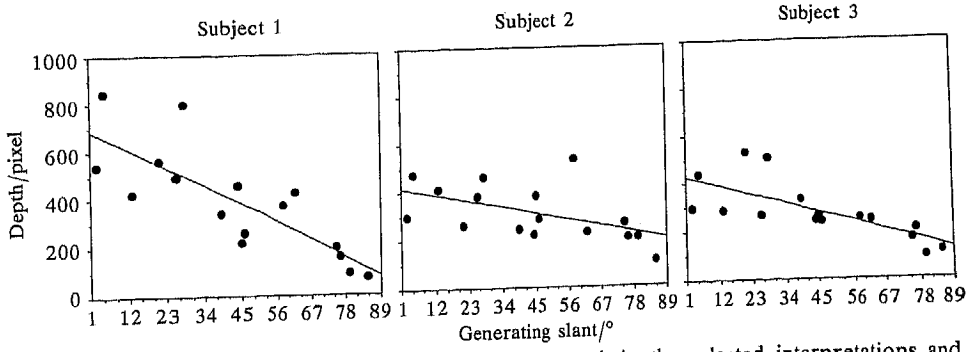

Figure 3. The correspondence between line-of-sight depth in the selected interpretations and generating slant in experiment 1.

perceived depth significantly. Research focusing on the relationship between projected width and perceived depth will be necessary to address this issue (eg Caudek and Proffitt 1993).

4 Experiment 2

The systematic relation between selected and generating interpretations implies that subjects use constraints other than rigidity to recover a unique interpretation in two-view displays. Experiment 2 addresses the question of whether constraints similar to those used to interpret two-view displays are also used when a rigidity constraint is theoretically sufficient to recover the 3-D structure. Under the assumptions of rigidity and orthographic projection, 3 or more views of at least four noncoplanar points are mathematically sufficient for recovery of the structure (the 3-D coordinates of the feature points) and motion (the rotation axis and rotation magnitude) of the object, up to a reflection in the image plane (Ullman 1979). In contrast, recent work by Todd and his colleagues (Norman and Todd 1993; Todd and Bressan 1990; Todd and Norman 1991) indicates that the additional information present in SFM displays with more than two views is not used by human subjects to determine object shape.

By using the 3-D generating objects of experiment 1, thirty-view displays were created which had the same terminal views as the two-view displays, plus 28 additional views in between. Subjects were asked to find the best match to these displays among the set of 89 response displays depicting interpretations compatible with the two terminal views. Note, however, that only one of the 89 response displays depicts an interpretation compatible with the thirty-view display, ie only one depicts the generating interpretation. If subjects do not use the additional information present in multiview displays to determine object shape, then we would not necessarily expect subjects to select the generating interpretations.

4.1 *Method*

4.1.1 *Subjects.* The subjects were two graduate students from the University of California, Irvine, who were paid for their participation. Neither subject had participated in experiment 1. Subject 1 was knowledgeable and subject 2 was naive as to the purposes in the experiment. Both had normal or corrected-to-normal vision.

4.1.2 *Design.* Two independent variables were examined: (1) generating slant, the same sixteen as in experiment 1, and (2) the number of views in the oscillating display, 2 or 30. The dependent variable was the slant index of the selected interpretation.

4.1.3 *Stimuli.* The same two-view displays and response displays (depicting the interpretations) were used as in experiment 1. 16 additional oscillating displays were created, each of which contained 30 views. For each two-view display a thirty-view display was created by adding 28 views in between the 2 views. The 3-D model which was used to generate the two-view display was simply rotated from frame 1 of the two-view sequence through 29 equal steps, 1 step for each view, until the second frame of the two-view display was reached. Clearly, any one of the infinite number of interpretations could have been used to 'fill in' the motion between the 2 views, but, given the results of experiment 1, we chose the generating interpretation. There are two important facts to note about our procedure for adding views. First, adding more views did not affect the display duration. The SOA between each of the 30 views was 13.3 ms so that the SOA between end frames was still 400 ms (as in experiment 1). A further benefit was that these SOAs were close to those which Todd et al (1988) found to maximize judgments of rigidity for two-view and multiview displays. Second, adding more views did not provide the subjects with more extreme views of the

simulated object. Todd and Norman (1991) have argued that all views of the same object do not provide the same amount of information about the shape of the object. They found that performance in discriminating spheres and ellipsoids was as good with 2 nonfrontal views as with 8 views but was not as good with 2 near-frontal views.

Each response display for a given stimulus display was transformed by a global 3-D rotation so that the axis of rotation always had the same direction. The procedure was equivalent to changing the subject's viewpoint for each interpretation so that the rotation axis was always in the same 3-D relation to the subject. The slant of the rotation axis was chosen randomly between 5° and 25° and was at least 9° from the generating slant. This prevented there being a direct 2-D match between frames in a two-view or thirty-view display and frames in its correct interpretation. (This control is only relevant for the thirty-view displays, for which the presence of matching frames, other than the first and last frames, could have revealed the generating interpretation.) The tilt of the rotation axis was chosen randomly between 0° and 180° and was also restricted from being within 9° of the generating tilt. It is possible that in experiment 1, response displays with rotation axes near the line of sight (ie near 89°) were perceived to have less depth than was simulated (Loomis and Eby 1988). Relocating the rotation axis such that it was near the image plane should produce more-veridical perceptions of depth. In addition, using the same rotation axis for all of the response displays eliminates the relationship between the slant index of the selected interpretation and the slant of the rotation axis. This procedure assures that variations in perceived depth in the interpretations were not due to systematic variations in their rotation axes.

4.1.4 *Procedure*. The procedure was essentially the same as that in experiment 1. The 32 displays (sixteen slants seen in 2 views or 30 views) were completely randomized and divided into two experimental sessions. This process was repeated four times, for a total of eight experimental sessions. There were eight practice trials in the first session, four in sessions two through four, and two in sessions five through eight. The practice displays were all different from the experimental displays but were generated in the same way. The practice trials included two-view and thirty-view displays.

4.2 *Results and discussion*

For each display the slant index of the selected interpretation was averaged across the four judgments and compared with the generating slant. The correspondences between the two are shown in figure 4. The result found for two-view displays in experiment 1 was replicated. There was a significant correlation ($p < 0.05$) between selected and generating interpretations ($r_{14} = 0.679$ for subject 1 and 0.793 for subject 2). The correlation was also significant for thirty-view displays ($r_{14} = 0.862$ and 0.915, respectively). As shown in figure 4, the slope of the regressions were similar for the two-view and thirty-view displays.

We also examined the relationship of the depth in the selected interpretations to the generating slant. As in experiment 1, the depth in the selected interpretations decreased as the generating slant approached the viewing direction. This was true for two-view displays ($r_{14} = -0.817$ and -0.539 for subjects 1 and 2, respectively) and for thirty-view displays ($r_{14} = -0.719$ and -0.620, respectively). The depth simulated in the selected interpretations was, on average, similar to the depth in the generating interpretation for subject 1 (4.8 pixels greater for two-view displays and 0.3 pixels greater for thirty-view displays) and less than the depth in the generating interpretation for subject 2 (280.2 pixels less for two-view displays and 289.8 pixels less for thirty-view displays). The relationships between generating slant and the depth in the selected interpretation are presented in figure 5.

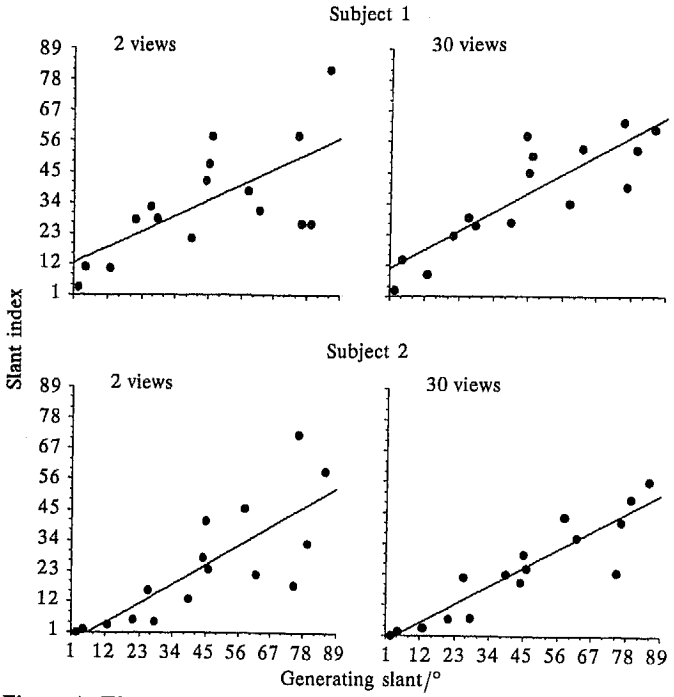

Figure 4. The correspondence between slant index of the selected interpretations and generating slant in experiment 2.

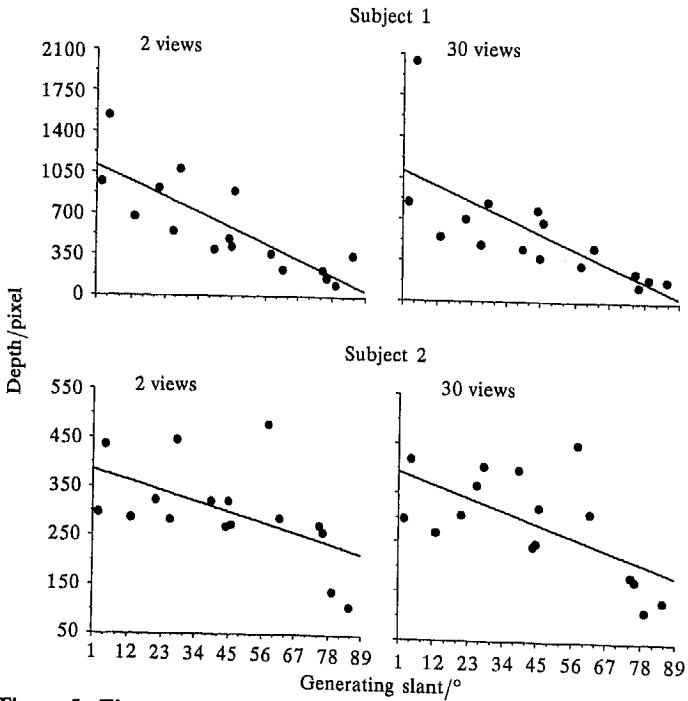

Figure 5. The correspondence between line-of-sight depth in the selected interpretations and generating slant in experiment 2.

According to Ullman's (1979) theorem, 3 views of four noncoplanar points are sufficient to recover accurately 3-D metric structure. Nevertheless, there was little change in the slope of the regression line relating the generating slant to the slant index of the selected interpretations when additional views were added. That the pattern of judgments was similar to that of experiment 1, in which only two-view displays were examined, supports the conclusion by Todd and his colleagues (Norman and Todd 1993; Todd and Bressan 1990; Todd and Norman 1991) that subjects do not make full use of the extra information available in multiview displays of this type to recover object shape. Instead, subjects may base their perception of structure on the same heuristic processes that are used to interpret two-view displays.

5 Experiment 3

The correspondence between selected and generating interpretations found in experiments 1 and 2 suggests that the constraints used to recover structure in two-view displays are related to the constraints which we used when generating the displays. For example, the points in the generating objects were confined to a sphere and the rotation (in degrees) between the extreme views was selected from the interval [8, 18]. In experiment 3 we systematically varied the shape of the generating objects and the 3-D rotation magnitude to determine whether the correspondence between the selected interpretations and the generating interpretations in the previous experiments was based on these constraints. We began with displays generated in the same manner as in the previous experiments, but added three additional versions of each display by increasing the depth in the 3-D model, decreasing the rotation speed, or both. The displays with increased depth, which we will refer to as 'stretched', had twice the extent in depth as the original version. The displays with reduced rotation speed rotated at 71.25% of the rotation speed of the original version. The fourth variant had both increased depth and reduced rotation speed.

5.1 *Method*

5.1.1 *Subjects.* The subjects were four graduate students from the University of California, Irvine, who were paid for their participation. None of the subjects had participated in the first two experiments and all were naive to the purposes of the experiment. All of the subjects had normal or corrected-to-normal vision.

5.1.2 *Design.* Four independent variables were examined in this experiment: (1) the generating slant (five levels); (2) the depth simulated in the generating object (unstretched, as in experiments 1 and 2, or stretched, with twice the line-of-sight depth); (3) the rotation angle between the extreme views in the generating object (large rotation, as in experiments 1 and 2, or small rotation, which was 71.25% of the large rotation); and (4) number of views (2 or 30). The dependent variable was the slant index of the selected interpretation.

5.1.3 *Stimuli.* Two sets of 8 displays were generated for each of the five levels of the generating slant index. The displays in each set had similar 2-D dot configurations but differed with respect to simulated depth, simulated rotation magnitude, and number of views. For each set of 8 displays we randomly selected five points within the volume of a sphere, as in experiments 1 and 2. These five points were used for the unstretched displays. To produce the stretched displays, the z coordinates of these points were scaled by a factor of 2. The x and y coordinates were left unchanged. These two sets of points were combined factorially with two rotation angles, ρ and ρ', to produce 4 displays. The rotation angle ρ was selected at random from the interval [8, 18] (as in the experiment 1 and 2 displays). ρ' was equal to 0.7125ρ. The factor 0.7125 was chosen to produce similar average 2-D dot

displacements for the two views of the unstretched large-rotation displays and the stretched small-rotation displays. (Since we were simulating slanted axis rotations, an exact match could not be produced.) The same rotation axis was used for all 4 displays. The slant of the rotation axis was selected at random from one of the eight 11° intervals described in experiment 1 and the tilt was chosen randomly in [0, 180]. Frame 1 of each display was produced by rotating one of the sets of points backwards by half of the selected rotation angle and frame 2 was produced by rotating the set of points forward by the same amount. The thirty-view displays were produced by rotating the set of points forward from frame 1 of the two-view display through 29 equal steps to frame 2. Eight sets of displays were generated in this manner, one set for each of the slant intervals. The point-selection criteria described in experiment 1 were applied to the unstretched large-rotation displays and the unstretched small-rotation displays. In addition, none of the dots in the projection of the stretched small-rotation displays were allowed to extend beyond the projection of the sphere bounding the unstretched displays. This prevented there being a simple 2-D size cue which would have identified the stretched displays. On average fifteen points were discarded while generating each of the eight display sets in order to meet these criteria.

It is important to note that the one-parameter families of interpretations for the two-view displays in each set of 4 displays (ie those displays having the same generating slant) are not the same. The families for the 2 unstretched displays share one common 3-D shape, ie the generating object, as do the families for the 2 stretched displays, but there are no common interpretations between the families of the stretched and the unstretched displays. All four families however, contain interpretations with similar line-of-sight depth.

One possible explanation for the similarity in judgments between the two-view and thirty-view displays in experiment 2 is that subjects chose an interpretation for a thirty-view display and then used that interpretation when making a selection for its two-view counterpart. Although this is an unlikely possibility given the similarity of the experiment 2 results to those of experiment 1 (in which there were no thirty-view displays), we decided to create two sets of displays at each generating slant, so that it would be impossible to make such comparisons. Each subject viewed displays from one set on two-view trials and displays from the other set on thirty-view trials. To make the alternative displays, we permuted the assignment of z coordinates to (x,y) pairs among the original set of five points and then followed the procedure described above to construct the related displays. We used the same rotation axis and rotation angles so that the generating slant and line-of-sight depth would be the same for corresponding displays in the two sets. The assignment of display set to number of views was counterbalanced across subjects.

As previously discussed, some of the points in some of the response displays in experiments 1 and 2 near the extremes of the 1° to 89° slant-index range did not remain within the limits of the display area for the entire 360° rotation cycle. Since the interpretations selected by the subjects in experiments 1 and 2 did not approach these limits, we elected to eliminate from the present experiment interpretations that did not remain within the limits of the display area. Recall that the direction of the rotation axis for the interpretations in experiment 2 was fixed on each trial. For the present experiment, the direction of this axis was selected (within the same range as in experiment 2) so as to maximize the range of interpretations remaining fully within the bounds of the display oscilloscope. Because the increase in size was typically concentrated in a single direction for the displays used in this series of experiments, and because our display oscilloscope was wider than it was high, we switched the side-by-side arrangement of the stimulus and interpretation to a top-bottom arrangement. Once the maximum range was determined, it was divided into 89 equally spaced

slant values. The range of slants was the same for all 8 displays in each display set. Of the eight levels of generating slant used in experiments 1–3, we excluded from this experiment the two levels with generating slants closest to the image plane and the level with generating slant closest to the line of sight because their range of usable interpretations (interpretations for which all points remained on the display oscilloscope on all frames) was too limited.

5.1.4 *Procedure.* Each subject participated individually in two 30 min practice sessions and eight 1 h experimental sessions. The sessions were blocked by number of views. The practice sessions consisted of eight trials presented in random order. One practice session contained two-view displays and the other contained thirty-view displays. Each experimental session consisted of two new practice trials followed by twenty experimental trials (5 generating slants × 2 simulated depths × 2 simulated rotation angles) presented in random order. The subjects were allowed to rest half way through each experimental session. The eight experimental sessions were run in the order AABBBBAA with respect to number of views for two subjects and BBAA-AABB for the other two subjects.

The instructions to the subjects were the same as in experiment 2. The subjects were required to view the full range of response displays before making a final selection.

5.2 *Results and discussion*

As indicated above, the interpretation families are different for two-view displays having the same generating slant but different simulated depths and simulated rotation magnitudes. This means that the rotation-axis slant of the selected interpretations cannot be meaningfully compared across these conditions. The same rotation-axis slant does not correspond to the same 3-D configuration. The depth in the selected interpretations, however, can be compared across conditions. To assess the effects of varying the depths and rotation magnitudes used to generate the displays, we conducted a four-way (5 generating slants × 2 simulated depths × 2 simulated rotation magnitudes × 2 numbers of views) repeated-measures analysis of variance (ANOVA) with line-of-sight depth in the selected interpretations as the dependent variable. The main effect of simulated depth in the generating object was significant ($F_{1,3} = 339.67$, $p < 0.05$, $\omega^2 = 0.197$) and indicated that subjects selected interpretations with more depth for the stretched displays. This effect can be seen in figure 6. The main effect of rotation magnitude in the generating object was also significant ($F_{1,3} = 10.66$, $p < 0.05$, $\omega^2 = 0.020$). The depth in the selected interpretations was greater for displays generated with larger rotation angles between end views. There was

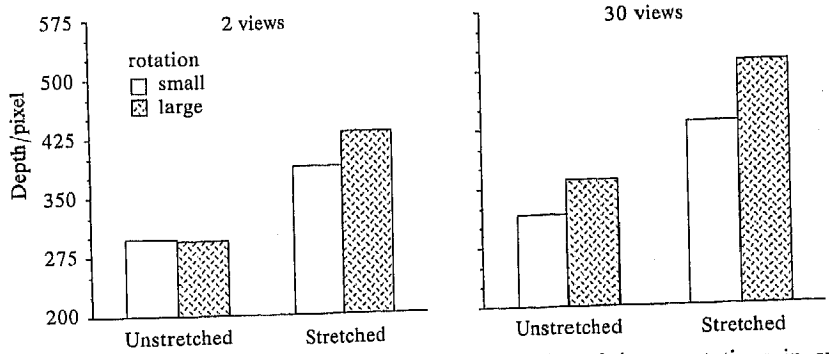

Figure 6. The average line-of-sight depth in the selected interpretations in experiment 3. In each panel are four conditions corresponding to two levels of simulated depth and two levels of simulated 3-D rotation.

a significant interaction between generating depth and generating rotation angle ($F_{1,3} = 35.03$, $p < 0.05$, $\omega^2 = 0.004$). The effect of generating rotation angle was greater for the stretched displays. The only other significant effect was an interaction between generating slant and simulated depth ($F_{4,12} = 11.94$, $p < 0.05$, $\omega^2 = 0.054$).

These results indicate that the selected interpretations were affected by manipulations of both the shape and the motion of the generating object. When we increased either the simulated depth or the simulated rotation between end views, the subjects selected interpretations having greater line-of-sight depth. This occurred both for multiview displays and for two-view displays. One might expect this result for stretched multiview displays under the assumption that subjects were accurately recovering the structure in these displays. However, even though subjects chose interpretations with greater line-of-sight depth for the stretched displays, they did not select the generating objects. The depth in their selections was always much less than what was simulated. The average depth simulated in the unstretched displays was 571 pixels whereas the average depth in the selected interpretations was 296 pixels for two-view displays and 339 pixels for thirty-view displays. For stretched displays, the average simulated depth was 1139 pixels and the average depth in the selected interpretations was 410 pixels (two-view) and 471 pixels (thirty-view). In addition to the nonveridical depth percepts, the effect of rotation magnitude on the depth in the selected interpretations cannot be accounted for by a veridical geometric analysis of the stimuli.

This experiment provides further evidence that similar constraints are used to interpret two-view and multiview displays and that the additional information present in multiview displays is not being used to recover veridical object shape. What information then is used to give a 3-D interpretation to these displays? Earlier we discussed two possible constraints which could have produced the correspondence found in experiments 1 and 2 between generating interpretations and the interpretations selected by the subjects: (1) subjects are biased toward interpretations which are no deeper than they are wide, and (2) subjects are biased toward interpretations simulating small rotations between end views. If subjects responded to these displays as if they were similar in depth, we would expect interpretations with similar depth to be selected regardless of the depth and rotation of the generating object, at least for the two-view displays for which depth is geometrically unspecified. Subjects did not select interpretations that had the same depth, however, as simulated depth and rotation magnitude varied. Instead, they selected interpretations with greater depth when either the depth or the rotation magnitude of the simulated object increased.

Both increased depth and increased rotation magnitude in the generating object increase relative image motion, and for this reason we considered the possibility that perceived depth in two-view displays and in multivew displays is a function of relative image motion. (Proffitt et al 1992 have recently presented a similar hypothesis in the context of stereokinetic displays.) To test this hypothesis we developed a measure of relative motion that could be applied to two-view displays or to the terminal views of multiview displays. First, consider the image motion resulting from a 3-D rotation. Any 3-D rotation can be expressed as a rotation about the line of sight followed by a rotation about an axis in the image plane (see, for example, Green 1959). Rotation about the line of sight produces only rotary motion or 'curl' in the image and will not affect the relative separation of dots in the image. Since this rotation does not contribute to relative motion, we eliminate the curl in our analysis by rotating the dots in frame 2 about the line of sight until the remaining motion trajectories in the image are parallel [see Todd and Bressan (1990) for a description of this procedure]. Relative motion is then defined as the maximum (signed) difference among the remaining displacements. Consider what happens to relative motion when the slant of the rotation axis is manipulated but the magnitude of the 3-D rotation is not.

When the rotation axis is in the image plane there is no curl component of motion and all of the image motion contributes to relative motion. When the rotation axis is slanted, relative motion will decrease because part of the overall image motion is curl. With the curl component removed, the magnitude of the remaining parallel displacements is diminished. Relative motion will be zero when the rotation axis is coincident with the line of sight. The left panel of figure 7 shows how relative motion changes when the same object is rotated about axes with different slants. The right panel of figure 7 shows the actual relative motion in the 16 displays in experiments 1 and 2. As the slant of the generating rotation axis approached the viewing direction, the magnitude of relative motion decreased. This is precisely what happened to the depth in the selected interpretations in experiments 1 and 2.

We computed the correlation between relative motion and depth in the selected interpretations. The correlation was significant ($p < 0.05$) for all three subjects in experiment 1 ($r_{14} = 0.814$, 0.522, and 0.696) and both subjects in experiment 2 [$r_{14} = 0.842$ (two-view) and 0.721 (thirty-view) for subject 1, and 0.537 and 0.636, respectively, for subject 2]. The depth variations in the selected interpretations in the present experiment were also related to variations in relative motion. The correlation (considered over all 20 displays—5 generating slants × 2 depths × 2 rotations) was significant ($p < 0.05$) for three of the four subjects ($r_{18} = 0.782$, 0.667, and 0.761 for two-view displays and 0.799, 0.789, and 0.747 for thirty-view displays). The correlations for subject 3 failed to reach significance ($r_{18} = 0.316$ for two-view displays and 0.397 for thirty-view displays). A more detailed analysis of subject 3's responses revealed that, within each set of four displays having the same generating slant but different simulated depth and rotation, his responses were ordered by relative motion for fourteen out of fifteen cases for thirty-view displays but only six out of fifteen for two-view displays. Figure 8 shows the depth in the selected interpretations compared with the relative motion in the displays. There is one data point for each of the four combinations of simulated depth and simulated rotation.

If relative motion was being used to interpret the structure in these displays, there is still the problem of determining what constraints might underlie this relationship. Relative motion is affected both by variations in relative depth within an object and by variations in rotation magnitude. One possibility is that a fixed amount of depth is perceived and relative motion leads to variations in perceived rotation magnitude. This possibility is not consistent with our results, which show an increase in the depth in the selected interpretation with increased relative motion. Another possibility is

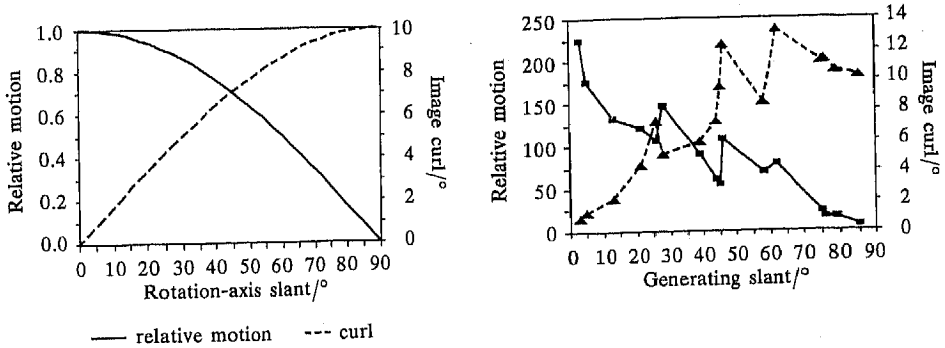

Figure 7. Image curl and relative motion (without curl) for an object rotating 10° about increasingly slanted axes (left panel); the same image measures for the 16 displays studied in experiments 1 and 2 (right panel). In the left panel relative motion is shown as a proportion of the relative motion resulting from rotation about an axis with 0° slant. In the right panel relative motion is shown in pixels, as measured in the actual displays.

that the perceived rotation magnitude is constant across displays and variations in relative motion lead to variations in perceived depth. This is consistent with the results of the present experiment, but in this experiment perceived depth and perceived rotation magnitude were not assessed directly. In experiment 4 we obtained a more direct judgment of perceived depth and in experiment 5 we obtained judgments of perceived rotation.

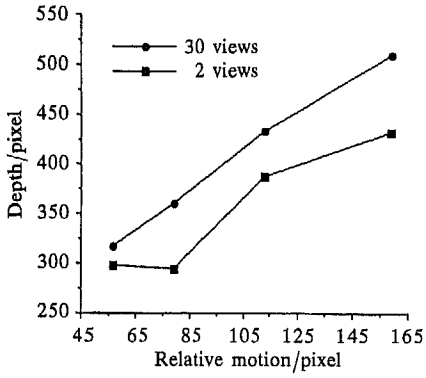

Figure 8. The relationship between line-of-sight depth in the selected interpretations and relative motion in experiment 3. The four data points for each number of views correspond (from left to right) to unstretched small-rotation displays, unstretched large-rotation displays, stretched small-rotation displays, and stretched large-rotation displays.

6 Experiment 4

Thus far, we have based our conclusions concerning perceived depth on the depth present in the interpretation which subjects chose as the best match to the structure perceived in an oscillating display. Although the results have been robust, this measure of perceived depth is an indirect one. In the present experiment, instead of choosing from a family of interpretations, subjects were asked to make a direct judgment of the depth in the oscillating displays. Subjects adjusted the separation of two dots which appeared on the bottom of the display oscilloscope to indicate the perceived depth in the oscillating display.

6.1 *Method*

6.1.1 *Subjects.* The subjects were Jeffrey Liter, plus three graduate students from the University of California, Irvine, who were paid for their participation. One of the graduate students had participated in experiment 2. The other subjects had not participated in any of the other experiments but were aware of the purposes of the experiment. All of the subjects had normal or corrected-to-normal vision.

6.1.2 *Design.* Four independent variables were examined: (1) generating slant (42° or 68°), (2) simulated depth in the generating object (unstretched or stretched), (3) simulated rotation magnitude (small or large), and (4) number of views (2 or 30). The dependent variable was the perceived line-of-sight depth and was measured by having the subjects adjust the horizontal separation of two dots which appeared on the bottom of the display oscilloscope.

6.1.3 *Apparatus.* The apparatus was the same as in experiments 1, 2, and 3 except that the computer which controlled the display oscilloscope was a Vax Station II. This apparatus was used for the remainder of the experiments.

6.1.4 *Stimuli.* Two sets of displays with intermediate generating slants, display set 2 (42°) and display set 5 (68°) from experiment 3, were selected as stimuli for the

current experiment. On each trial, a display from one of these sets was presented in the middle of the display oscilloscope and two dots, aligned horizontally, were presented on the bottom of the display oscilloscope. Using a joystick, the subject could adjust the horizontal separation of the two dots in real time as the display oscillated. The range of separation was 0 pixels to 2048 pixels in increments of 2 pixels. The initial separation of the dots was random within this range.

6.1.5 *Procedure.* Each subject participated in three sessions. The first session was a practice session and the second two were experimental sessions. Each session was divided into two blocks, one for two-view displays and the other for thirty-view displays. The four experimental blocks were ordered ABBA with respect to number of views for two subjects and BAAB for the other two subjects. Each block consisted of sixteen trials—two replications of each of the 8 different displays (2 generating slants × 2 simulated depths × 2 simulated rotation magnitudes). 16 different displays were used for the practice blocks.

The subjects were informed that they were to judge the magnitude of depth along the line of sight by adjusting the separation of the two dots which appeared on the bottom of the screen. The subjects were told to imagine the object in a position halfway between the two extreme positions of the motion sequence and to judge the depth of the object on the basis of this intermediate position. They were shown a top-view representation of a five-dot display on a piece of paper to emphasize that they were to judge the separation along the line of sight and not along any other direction.

6.2 *Results and discussion*

The mean judged depths of the four replications for each display were computed for each subject and compared in a four-way repeated-measures ANOVA to determine the effects of generating slant, simulated depth, simulated rotation magnitude, and number of views. Stretched displays were judged to be deeper ($F_{1,3} = 45.58$, $p < 0.05$, $\omega^2 = 0.016$), but an interaction with number of views ($F_{1,3} = 20.79$, $p < 0.05$, $\omega^2 = 0.009$) indicated that this effect was based primarily on the thirty-view displays. The effects of simulated depth and simulated rotation angle on judged depth in two-view and thirty-view displays are shown in figure 9. Displays which simulated more rotation between the end views were also judged to be deeper ($F_{1,3} = 20.93$, $p < 0.05$, $\omega^2 = 0.006$). Although the interaction between simulated rotation angle and number of views was not significant ($F_{1,3} = 3.14$, $p > 0.05$), the rotation effect also appears to be based primarily on the thirty-view displays (see figure 9). Thirty-view displays were judged to be deeper than two-view displays ($F_{1,3} = 13.68$, $p < 0.05$,

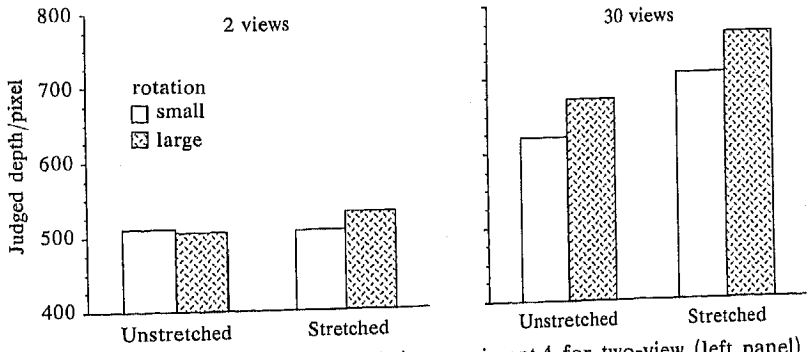

Figure 9. The average judged depth in experiment 4 for two-view (left panel) and thirty-view (right panel) displays. In each panel are four conditions corresponding to two levels of simulated depth and two levels of simulated 3-D rotation.

$\omega^2 = 0.193$). The displays in set 5 were judged to be deeper than those in set 2 ($F_{1,3} = 14.68$, $p < 0.05$, $\omega^2 = 0.074$). This ordering is in accord with the simulated depth in the generating objects. The average line-of-sight depth in the set 5 displays was 1073 pixels whereas that in the set 2 displays was 645 pixels. There were significant three-way interactions between simulated depth, simulated rotation angle, and number of views ($F_{1,3} = 87.41$, $p < 0.05$, $\omega^2 < 0.001$) and between simulated depth, simulated rotation angle, and generating slant ($F_{1,3} = 13.89$, $p < 0.05$, $\omega^2 = 0.003$). No other interactions were significant.

The significant effects of simulated depth and simulated rotation angle on judged depth are consistent with the hypothesis that perceived depth increases with relative motion. As shown in figure 10, judged depth increased with greater 2-D relative motion, regardless of its source—increased simulated depth or increased simulated rotation. The source of the relative-motion increase is in principle recoverable for a SFM display with more than two views, but the results for the thirty-view displays indicate that subjects failed to distinguish the source. Although the two data points on the left in figure 10 represent displays with the same simulated depth, judged depth increased with rotation angle. The same is true of the two points on the right. For two-view displays, judgments of depth were nearly constant across the different conditions. Subjects' subjective reports were that the judgment was more difficult for two-view displays.

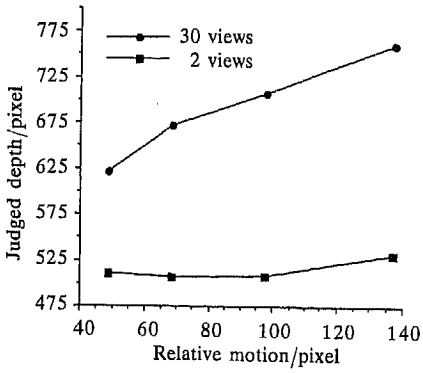

Figure 10. The relationship between judged depth and relative motion in experiment 4. The four data points for each number of views correspond (from left to right) to unstretched small-rotation displays, unstretched large-rotation displays, stretched small-rotation displays, and stretched large-rotation displays.

7 Experiment 5

The results of experiment 4 indicate that, at least for thirty-view displays, variations in relative motion result in variations in judged relative depth, regardless of the actual source of the relative motion variation—variations in relative depth or variations in rotation magnitude. One possible basis for this result is that subjects assume a constant amount of rotation and attribute variations in relative motion to variations in relative depth. This hypothesis is tested in experiment 5 by obtaining judgments of perceived rotation for the stimuli used in experiment 4.

7.1 *Method*

7.1.1 *Subjects.* The subjects were Jeffrey Liter, plus three graduate students from the University of California, Irvine, who were paid for their participation. Three of the subjects, including the author, had participated in experiment 4. The fourth subject had not participated in any of the previous experiments and was naive to the purposes in the experiment. All of the subjects had normal or corrected-to-normal vision.

7.1.2 *Design*. The first four independent variables were the same as in experiment 4: (1) the generating slant, (2) simulated depth in the generating object, (3) simulated rotation in the generating object, and (4) number of views. A fifth independent variable was added—the diameter of an oscillating hemisphere which subjects adjusted to indicate perceived rotation (230 or 460 pixels). The dependent variable was the judged rotation angle between the two end views in the five-point displays and was measured by having subjects adjust the rotation magnitude of a hemisphere which oscillated in phase with the five-point display.

7.1.3 *Stimuli*. The same stimulus displays were used as in experiment 4. On each trial, a stimulus display appeared on one side of the display oscilloscope, and an oscillating hemisphere appeared on the opposite side. The surface of the hemisphere was covered with 100 randomly placed points. We selected a hemisphere rather than a transparent sphere to minimize orthographic reversals, which we had found in preliminary observations with a transparent sphere. Subjects reported a strong bias to see the hemisphere as convex and did not have difficulty with orthographic reversals. The hemisphere oscillated in phase with the stimulus display, but its amplitude of rotation was under the control of the subject. By pushing forward or pulling backward on a joystick the subject could increase or decrease the rotation amplitude of the hemisphere in real time from 0° to 180° in 1° increments. The bounding (ie occluding) contour of the response hemisphere changed as the hemisphere rotated. In the frontal orientation its projected contour was a disk. When rotated 90° from frontal, its contour was like that of a half moon. The initial rotation amplitude was chosen randomly on each trial. The axis of rotation of the hemisphere (horizontal or vertical) and the side of the screen on which it appeared was counterbalanced across subjects.

As indicated above, two hemisphere sizes were used, with radii of 230 and 460 pixels. The radius of the hemisphere determines the linear velocities in the projection, for a given simulated 3-D rotational velocity. Although researchers have found only small effects of linear velocity on perceived 3-D rotational velocity (see, for example, Kaiser 1990; Kaiser and Calderone 1991; Petersik 1991), we thought it advisable to determine, for our displays, whether the linear velocity of the hemisphere would affect judgments of rotation magnitude of the five-point displays.

7.1.4 *Procedure*. Each subject participated in five sessions. The first session was a practice session and the following four were experimental sessions. The radius of the response hemisphere was the same for all of the displays within a session but varied across sessions. The sessions were ordered ABBA with respect to radius for two subjects and BAAB for the other two subjects. The size of the response hemisphere in the practice session was the same as it was in the first experimental session. Each session consisted of two blocks, one for two-view displays and one for thirty-view displays. The order of blocks with respect to number of views across the four experimental sessions was AB-BA-BA-AB for two subjects and BA-AB-AB-BA for the other two subjects. Each block contained sixteen trials (two repetitions of each of the 8 displays—2 generating slants × 2 simulated depths × 2 simulated rotations).

The subjects were informed that they were to judge the magnitude of rotation between the two extreme views in the oscillating five-point display by adjusting the angle through which the hemisphere rotated. They were informed that they should ignore the direction of the axis of rotation when making their judgments and to consider only the magnitude of rotation.

7.2 *Results and discussion*

For each subject the mean judged rotation angle was computed for each condition by averaging the data from the four replications. The means were then compared in a five-way repeated-measures ANOVA to determine the effects of generating slant, simulated depth, simulated rotation angle, number of views, and diameter of the response hemisphere. Displays which simulated greater rotation were judged to rotate more ($F_{1,3} = 48.07$, $p < 0.05$, $\omega^2 = 0.101$). This result was obtained both for two-view and for thirty-view displays. Similarly, displays which simulated more depth were judged to rotate more ($F_{1,3} = 29.62$, $p < 0.05$, $\omega^2 = 0.066$) both for two-view and for thirty-view displays. These effects are shown in figure 11. The only other significant effect was an interaction between the diameter of the response hemisphere and the generating slant ($F_{1,3} = 25.30$, $p < 0.05$, $\omega^2 = 0.001$). The main effect of hemisphere diameter was not significant. The mean judged rotation magnitude was similar for both response hemispheres (22.0° for the small hemisphere and 20.7° for the large hemisphere). This result is consistent with the findings of Kaiser (1990; Kaiser and Calderone 1991) and Petersik (1991) and suggests that subjects were able to disregard variations in linear velocity related to variations in hemisphere size when making rotation judgments for the five-dot displays.

The hypothesis that perceived rotation in these displays is constant across variations in simulated depth and rotation angle was disconfirmed. Judged rotation increased

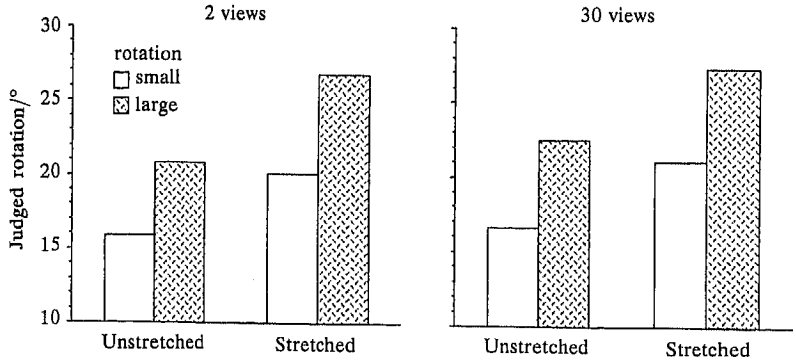

Figure 11. The average judged rotation in experiment 5 for two-view (left panel) and thirty-view (right panel) displays. In each panel are four conditions corresponding to two levels of simulated depth and two levels of simulated 3-D rotation.

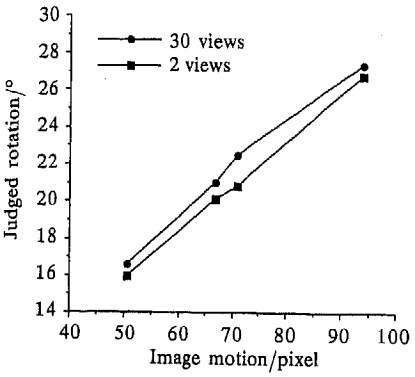

Figure 12. The relationship between judged rotation and average 2-D motion (with curl) in experiment 5. The four data points for each number of views correspond (from left to right) to unstretched small-rotation displays, stretched small-rotation displays, unstretched large-rotation displays, and stretched large-rotation displays.

both with greater simulated depth and with greater simulated rotation angle. A close examination of figure 11 suggests that judged rotation in these five-point displays is related to the average projected velocity. Recall that displays in the unstretched large-rotation condition and displays in the stretched small-rotation condition had similar average velocities. The rotation judgments for these displays were similar. As shown in figure 12, rotation judgments correspond quite well to average projected velocity.

8 General discussion

A rigidity constraint is not sufficient by itself to recover a unique 3-D structure from a two-view orthographic display. Therefore, our initial objective was to determine what additional constraints human vision might use. The results of experiment 1 indicate that the interpretations selected by subjects for two-view displays are systematically related to the 3-D structures that we used to generate the displays. In experiment 2 we found that subjects, when shown multiview displays, for which unique interpretations (plus reflections) can in principle be recovered, give 3-D interpretations similar to those given to two-view displays constructed from the first and last views of the multiview displays. Although surprising from a computational vision perspective, this result supports the conclusions of Todd and his collaborators (Norman and Todd 1993; Todd and Bressan 1990; Todd and Norman 1991): multiview displays do not, in general, provide more accurate recovery of 3-D structure than do two-view displays. We found little change in the function relating the selected interpretations to the generating interpretations when we added additional views.

Although the results of experiments 2 and 3 showed that similar 3-D interpretations were selected for two-view and multiview displays, the results of experiment 4 indicated a difference in the amount of depth reported for these types of displays. Multiview displays were judged to have greater depth. Subjective reports uniformly indicated that multiview displays appeared more three-dimensional. Some observers had difficulty perceiving depth in two-view displays but readily perceived depth in multiview displays. This difference in the saliency of the 3-D percept in two-view and multiview displays may have affected the direct judgments of depth in experiment 4 more than it affected the selection of matching structures in experiments 1 and 2. Additional views clearly have a perceptual effect, even if they do not provide accurate recovery of the 3-D structure.

The results of experiments 1, 2, and 3 suggest that subjects, in interpreting two-view and multiview displays, use constraints that are similar to those we used to generate the displays. Since the displays were generated within a spherical volume and with a limited range of rotation angles, a preferred perceived depth or a preferred perceived rotation angle, similar to the depths and rotation angles used to generate the displays, could account for these results. If subjects prefer to see a particular depth in a display, a depth perhaps related to the display width (see Caudek and Proffitt 1993), then we would expect 3-D interpretations with similar depths to be selected for two-view displays that were similar in projected width, regardless of the depth or rotation angle of the generating object. If subjects prefer to see a particular rotation angle for all displays, we would expect variations in relative motion across displays to be perceived as variations in depth. A deeper object would be perceived whether the increased relative motion was produced by increasing the simulated depth or increasing the simulated rotation angle. In experiment 3 we found that increasing the depth of the generating object or increasing the rotation angle led to an interpretation with greater depth. This result was confirmed with direct depth judgments in experiment 4, although the effects were mainly due to the multiview displays.

The increase in perceived depth with increased simulated depth or increased simulated rotation is consistent with a hypothesis of a constant perceived extent of rotation, with variations in simulated rotation being interpreted as variations in depth. This hypothesis, however, was not supported by the results of experiment 5, in which we obtained direct judgments of magnitude of rotation. Instead we found that either increasing the simulated depth or increasing the simulated rotation angle increased the magnitude of judged rotation. An alternative hypothesis, consistent with the results of experiments 3, 4, and 5 is as follows. (1) The human visual system extracts components of the image velocities that are not due to curl. We will not speculate at this time on how this is done. It may well be an approximate rather than a precise mathematical removal of the curl component. Removal of the curl component has been proposed in theories of recovery of heading from optic flow (for example, Koenderink and van Doorn 1981; Longuet-Higgins and Prazdny 1980; Perrone 1992) and is compatible with the vector-analysis approach discussed by Börjesson and von Hofsten (1975; see also Börjesson and Ahlström 1993). (2) With curl removed, the pair of points that has the maximum signed difference in velocity is selected from the $\binom{n}{2}$ possible pairs of points. This could be accomplished by comparing the points with the maximum velocities in each direction, or comparing the points with the minimum and maximum velocities if all points are moving in the same direction. (3) A measure of perceived depth in the display is derived from this maximum signed difference between image velocities. The perceived depth may be affected by other image variables such as display width (Caudek and Proffitt 1993; Proffitt et al 1992) and compression (Braunstein et al 1993). Exactly how the measure of perceived depth is scaled by other image variables is an important issue requiring further research.

Our results are consistent with Proffitt et al's (1992) emphasis on relative motion as a heuristic for determining perceived depth. They proposed that object-relative depth in kinetic-depth and stereokinetic cones is "strongly related to the magnitude of eccentricity (e/r)" (Proffitt et al 1992, page 20), where e is the projected displacement of the tip of the cone with respect to the center of the base and r is the radius of the base of the cone (ie one half the projected width of the display). The projected displacement of the tip after a complete oscillation cycle (ie $2e$) corresponds to the measure of relative motion discussed in the preceding paragraph since their rotations produced no image curl and the minimum image displacement was zero (the base of the cone was stationary). Proffitt et al divided the displacement by the radius to account for judgments of depth-to-width ratio for cones varying in radius.

The relative-motion hypothesis presented here can also account for the results that Loomis and Eby (1988) obtained with rotating SFM spheres. They found that judged depth was greatest when the axis of rotation was in the image plane and decreased as the slant of the rotation axis approached the line of sight. As in the simulation presented in the left panel of figure 7, the relative motion (per unit time) in their displays decreased as the slant of the rotation axis approached the line of sight. Loomis and Eby (1989) discuss two possible interpretations of their results: (1) human vision uses an algorithm capable of computing a veridical interpretation, but fails to do so for slanted axes because biological limitations make the required information unavailable or unreliable, or (2) human vision uses heuristic processes that provide only an approximate solution to the recovery of shape from motion, with the approximation becoming less veridical with increasing axis slant. Our results favor the second interpretation: relative motion, excluding curl, is used heuristically to specify perceived relative depth in SFM displays.

If perceptual processes are useful because they exploit environmental regularities, we might ask what regularities relate relative velocity to perceived depth. There are two situations in which the correlation between relative velocity in a 2-D image and relative depth in a 3-D scene is especially high. In any perspective translation, the relative depth of any two points that momentarily occupy the same image coordinates is given by the ratio of their image velocities (Longuet-Higgins and Prazdny 1980). (In the present situation, orthographic projections of rotation, we cannot convert velocity differences into velocity ratios in a meaningful way because we typically have near and far points moving in opposite directions. We could convert to ratios by assuming fixation on an intermediate point, but our choice of the fixation point would be arbitrary. However, differences can in general be converted to ratios if the mean is known.) The second situation in which relative velocities are highly correlated with relative depths is rotation about an axis in the image plane. If we consider only effects of orthographic projection (perspective effects are handled in the first situation), relative velocity is a function of relative depth and rotation magnitude. If an observer fixates a stationary object while translating (eg walking past the object) a rotation of the scene about an axis roughly in the image plane is projected onto the retina (see Hoffman and Bennett 1986). We can thus speculate that perceptual processes that relate relative velocity to relative depth are useful because they can be applied to these two common situations: perspective translations and rotations about axes in (or near) the image plane. We should emphasize that we are not suggesting that an observer misperceives rotation about a slanted axis, only that there is an automatically applied perceptual process that provides less-veridical judgments as the axis is moved out of the image plane.

Our primary concern in the present research was with perception of 3-D structure. Factors underlying amount of perceived rotation require further investigation. On geometrical grounds, we might expect a trade-off between perceived depth and amount of perceived rotation about an axis in the image plane, for a fixed amount of relative motion. We found no evidence of such a trade-off, even in an experiment (not included in this report) in which depth and rotation judgments were made on the same trial. The results of experiment 4 indicate that perceived depth is a function of relative motion in the image, without perceived rotation being taken into account (as judged in experiment 5). This differs from results obtained with oscillating dihedral angles. Braunstein et al (1993) found that the relationship between relative motion and judgments of relative depth depended on information for rotation, specifically compression of the projection of the object in the dimension perpendicular to the axis of rotation. A likely reason for this discrepancy is that the densely textured, planar surfaces studied by Braunstein et al (1993) provide richer information for judging rotation than the five-dot displays in the present study. Also, the investigation with dihedral angles used rotation about a vertical axis, whereas the present stimuli involved slanted rotation axes. Processes that recover depth from rotation may use amount of rotation more effectively when the axis is in the image plane, although this is yet to be determined. The relationship of average velocity to rotation judgments, found in experiment 5, may also be related to our use of very-low-density textures. There have been several interesting studies recently of the perception of rotation speed in 2-D displays (eg Werkhoven and Koenderink 1991) and in 3-D objects with planar facets or high texture densities (eg Kaiser 1990; Kaiser and Calderone 1991; Petersik 1991). Extension of this research to sparse textures and slanted rotation axes should prove useful.

Acknowledgements. This research was supported by National Science Foundation Grants IRI-8700924 and DBS-9209773 and Office of Naval Research Contract N00014-88-K-0354. The results of experiments 1-3 were presented at the annual meeting of The Association for Research in Vision and Ophthalmology, Sarasota, FL, May 1991, and the results of experiments 4 and 5 were presented at the same meeting in May 1992. We would like to thank Sergio Fajardo and Jill Nicola for their contributions at an early stage in this research, and Marc Albert, George J Andersen, Bruce Bennett, Asad Saidpour, and Jessica Turner for helpful discussions.

References

Aloimonos J, Brown C M, 1989 "On the kinetic depth effect" *Biological Cybernetics* **60** 445-455

Bennett B M, Hoffman D D, 1985 "The computation of structure from fixed-axis motion: nonrigid structures" *Biological Cybernetics* **51** 293-300

Bennett B M, Hoffman D D, Nicola J E, Prakash C, 1989 "Structure from two orthographic views of rigid motion" *Journal of the Optical Society of America A* **6** 1052-1069

Börjesson E, Ahlström U, 1993 "Motion structure in five-dot patterns as a determinant of perceptual grouping" *Perception & Psychophysics* **53** 2-12

Börjesson E, Hofsten C von, 1975 "A vector model for perceived object rotation and translation in space" *Psychological Research* **38** 209-230

Braunstein M L, 1976 *Depth Perception through Motion* (New York: Academic Press)

Braunstein M L, 1993 "Decoding principles, heuristics, and inference in visual perception", in *Perceiving Events and Objects: A Review of Gunnar Johansson's Research with Commentaries* Eds G Jansson, S S Bergstrom (Hillsdale, NJ: Erlbaum; in press)

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

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

Braunstein M L, Liter J C, Tittle J S, 1993 "Recovering 3-D shape from perspective translations and orthographic rotations" *Journal of Experimental Psychology: Human Perception and Performance* **19** 598-614

Caudek C, Proffitt D R, 1993 "Depth perception in motion parallax and stereokinesis" *Journal of Experimental Psychology: Human Perception and Performance* **19** 32-47

Doner J, Lappin J S, Perfetto G, 1984 "Detection of three-dimensional structure in moving optical patterns" *Journal of Experimental Psychology: Human Perception and Performance* **10** 1-11

Green B F Jr, 1959 "Mathematical notes on 3-D rotations, 2-D perspective transformations, and dot configurations" Group Report number 58-5 (Lexington, MA: Massachusetts Institute of Technology, Lincoln Laboratory)

Hoffman D D, 1982 "Inferring local surface orientation from motion fields" *Journal of the Optical Society of America* **72** 888-892

Hoffman D D, Bennett B M, 1985 "Inferring the relative 3-D positions of two moving points" *Journal of the Optical Society of America A* **2** 350-353

Hoffman D D, Bennett B M, 1986 "The computation of structure from fixed-axis motion: rigid structures" *Biological Cybernetics* **54** 71-83

Hoffman D D, Flinchbaugh B E, 1982 "The interpretation of biological motion" *Biological Cybernetics* **42** 195-204

Huang T, Lee C, 1989 "Motion and structure from orthographic projections" *IEEE Transactions on Pattern Analysis and Machine Intelligence* **11** 536-540

Johansson G, 1970 "On theories for visual space perception: A letter to Gibson" *Scandinavian Journal of Psychology* **11** 67-74

Kaiser M K, 1990 "Angular velocity discrimination" *Perception & Psychophysics* **47** 149-156

Kaiser M K, Calderone J B, 1991 "Factors influencing perceived angular velocity" *Perception & Psychophysics* **50** 428-434

Koenderink J, Doorn A J van, 1981 "Exterospecific component of the motion parallax field" *Journal of the Optical Society of America* **71** 953-957

Koenderink J J, Doorn A J van, 1991 "Affine structure from motion" *Journal of the Optical Society of America A* **8** 377-385

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, Prazdny K, 1980 "The interpretation of a moving retinal image" *Proceedings of the Royal Society of London, Series B* **208** 385 - 397

Loomis J M, Eby D W, 1988 "Perceiving structure from motion: Failure of shape constancy" in *Proceedings of the Second International Conference on Computer Vision* (Washington, DC: IEEE) pp 383 - 391

Loomis J M, Eby D W, 1989 "Relative motion parallax and the perception of structure from motion", in *Proceedings of the IEEE Workshop on Visual Motion* (Washington, DC: IEEE) pp 204 - 211

Norman J F, Todd J T, 1993 "The perceptual analysis of structure from motion for rotating objects undergoing affine stretching transformations" *Perception & Psychophysics* **53** 279 - 291

Perrone J A, 1992 "Model for the computation of self-motion in biological systems" *Journal of the Optical Society of America A* **9** 177 - 194

Petersik J T, 1991 "Perception of three-dimensional angular rotation" *Perception & Psychophysics* **50** 465 - 474

Proffitt D R, Rock I, Hecht H, Schubert J, 1992 "Stereokinetic effect and its relation to the kinetic depth effect" *Journal of Experimental Psychology: Human Perception and Performance* **18** 3 - 21

Todd J T, Bressan P, 1990 "The perception of 3-dimensional affine structure from minimal apparent motion sequences" *Perception & Psychophysics* **48** 419 - 430

Todd J T, Norman J F, 1991 "The visual perception of smoothly curved surfaces from minimal apparent motion sequences" *Perception & Psychophysics* **50** 509 - 523

Todd J T, 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, 1979 *The Interpretation of Visual Motion* (Cambridge, MA: MIT Press)

Ullman S, 1983 "Recent computational studies in the interpretation of structure from motion", in *Human and Machine Vision* Eds J Beck, B Hope, A Rosenfeld (New York: Academic Press) pp 459 - 480

Wallach H, O'Connell D N, 1953 "The kinetic depth effect" *Journal of Experimental Psychology* **45** 205 - 217

Werkhoven P, Koenderink J J, 1991 "Visual processing of rotary motion" *Perception & Psychophysics* **49** 73 - 82