

The spatial and temporal characteristics of perceiving 3-D structure from motion

DAVID W. EBY

University of California, Santa Barbara, California

In four experiments, a scalar judgment of perceived depth was used to examine the spatial and temporal characteristics of the perceptual buildup of three-dimensional (3-D) structure from optical motion as a function of the depth in the simulated object, the speed of motion, the number of elements defining the object, the smoothness of the optic flow field, and the type of motion. In most of the experiments, the objects were polar projections of simulated half-ellipsoids undergoing a curvilinear translation about the screen center. It was found that the buildup of 3-D structure was: (1) jointly dependent on the speed at which an object moved and on the range through which the object moved; (2) more rapid for deep simulated objects than for shallow objects; (3) unaffected by the number of points defining the object, including the maximum apparent depth within each simulated object-depth condition; (4) not disrupted by nonsmooth optic flow fields; and (5) more rapid for rotating objects than for curvilinearly translating objects.

The human visual system has the remarkable ability to recover three-dimensional (3-D) shape when it is presented with a rapid succession of 2-D views of a moving object. Even when each view by itself contains no information about three-dimensionality, depth can still be perceived. In their now classic study, Wallach and O'Connell (1953) named this phenomenon the *kinetic depth effect* (KDE). Recent investigators have called the phenomenon the recovery of *structure from motion* (SFM) (e.g., Todd, 1984; Ullman, 1979, 1984).

When viewing a KDE display, one often has the impression that the time course for the structural buildup is quite short. Wallach and O'Connell (1953) took note of this fact when they observed that "turning wire-figures were seen three dimensionally immediately upon presentation" (p. 208). Surprisingly, until recent years, there was very little data about the temporal characteristics of the process involved in the recovery of SFM (see Hildreth, Grzywacz, Adelson, & Inada, 1990;¹ Loomis & Eby, 1988; Todd & Bressan, 1990).

This lack of data is surprising for a number of reasons. First, because a static view of the 2-D imagery produces

no impression of depth (if all other cues to depth are removed), it should be obvious, once motion is initiated, that structural buildup is a fundamental aspect of the perception of SFM. Second, from an applied standpoint, data about the human ability to process rapid, kinetic 3-D information is often useful in system design. Data about the spatial and temporal characteristics of perceiving 3-D SFM could be used to optimize systems that rapidly display 3-D information, such as flight simulation systems. Third, as noted by several authors, information about how depth builds up over time is important for theories of the human perception of SFM (Grzywacz & Hildreth, 1987; Hildreth & Koch, 1987; Landy, 1987; Landy, Doshier, Sperling, & Perkins, 1988). My primary purpose in the present article is to provide an empirical background of temporal and spatial characteristics of the perception of SFM, by investigating the effects of several factors on the buildup of 3-D structure.

Several terms that I will use in this article are defined in Table 1. The buildup of 3-D structure can be defined either temporally or spatially. Since these two factors often covary (i.e., an increase in one often produces an equal increase in the other), they are frequently used interchangeably. However, in this and in other studies (e.g., Hildreth et al., 1990; Todd & Bressan, 1990), there is evidence to suggest that temporal and spatial variables affect the buildup of 3-D structure differentially. The buildup of 3-D structure is therefore defined as the function that relates perceived depth to *stimulus duration*; in spatial terms, buildup is defined as the function that relates perceived depth to *range of motion*. Most of the results in this study can be adequately described by exponential functions of two parameters—the asymptotic extension in depth and either the space constant (Equation 1) or the time constant (Equation 2). Thus, the buildup of 3-D structure is defined in relation to these scalar parameters.

A preliminary report of Experiments 1 and 2 was presented at the meeting of the Association for Research in Vision and Ophthalmology in Sarasota, FL in 1990. This work was supported by NSF Grant BNS-8919383 and grants from the Academic Senate of the University of California, Santa Barbara. For their comments on previous drafts of this article, I thank Jack Loomis, John Foley, Mike Braunstein, John Andersen, Jim Todd, Susan Ganter, and an anonymous reviewer. I thank Brian M'Closkey for his assistance with the statistical analyses. This work was completed in partial satisfaction of the requirements for the doctoral degree from the University of California, Santa Barbara. Correspondence may be addressed to D. W. Eby, Department of Cognitive Sciences, School of Social Sciences, University of California, Irvine, CA 92717.

Table 1
Definitions of Terms

Term	Definition
Angular increment	The number of degrees of curvilinear translation between each distinct view of an object and the next.
Distinct view	New information about an object consisting of a set of new 2-D coordinates.
Range of motion	The total range (either translation or rotation) through which an object moves during an entire display (i.e., the number of distinct views multiplied by either the rotary or the angular increment).
Refresh rate	The rate at which the video monitor redraws the information being displayed, independently of update rate, expressed in number of refreshes per second (in hertz).
Rotary increment	The number of degrees of rotation between each distinct view of an object.
Update rate	The number of distinct views of the object per second (in hertz).

In a study of the buildup of 3-D SFM, Hildreth et al. (1990) investigated three main variables: range of rotation (or display duration, since they covaried), simulated depth separation between the three points used in their display, and the effect of increased noise in the 2-D location of the image points. They found that the accuracy with which a subject could determine which of three points was intermediate in depth increased with increasing range of rotation up to about 30° or 40° and then leveled off. At their refresh rate of 33 Hz, this angular range corresponded to a duration of 660–900 msec. Using scalar judgments of perceived depth and curvilinearly translating objects, Loomis and Eby (1988) have studied the same variables and have found similar results, with reports of depth leveling off at about 40°–60° of motion. This range of motion corresponded to a duration of about 570–860 msec.

When different depths were simulated, Hildreth et al. (1990) found that performance accuracy increased as depth separation increased, suggesting that larger simulated separations were perceived as being deeper. This interpretation is supported by several studies in which it has been shown that subjects tend to report increased apparent depth when objects are simulated as being deeper (e.g., Eby & Loomis, 1989; Loomis & Eby, 1987a, 1988, 1989, 1990). In addition, Hildreth et al. (1990) found that, with large simulated separations, subjects' accuracy in indicating which of three points was intermediate in depth reached asymptotic levels at about the same time as it did with small simulated separations (however, the asymptotic levels differed, depending on the separation). A similar result has also been reported by Loomis and Eby (1988), who found that judgments of object depth for different-sized objects undergoing a curvilinear translation reached a maximum at about the same time.

The amount of noise in the 2-D positioning of points in a SFM display does not seem to affect the buildup of perceived structure. Hildreth et al. (1990) added noise to

the 2-D location of the points by randomly perturbing the *x* and *y* values of the points according to a Gaussian distribution. In different experimental sessions, the space constant for the Gaussian was varied to manipulate the amount of noise. Hildreth et al. found that as noise was increased, overall performance decreased; however, asymptotic levels of performance were reached at about the same angular range of rotation as in the condition in which no noise was present. This finding suggests that the addition of noise reduces the amount of depth separation perceived in the points but does not seem to affect the temporal or spatial characteristics of the build-up of structure. The results of the Hildreth et al. (1990) and Loomis and Eby (1988) studies provide some preliminary information about how object depth builds up.

The present study was designed as a systematic investigation of several other variables that are likely to be useful for theorizing about the process of recovering 3-D SFM. Specifically, the temporal and spatial characteristics of the buildup of perceived 3-D structure were investigated as a function of the speed of the object motion, the simulated object depth, the amount of surface overlap in transparent objects, the number of elements defining the object, and the type of simulated 3-D motion.

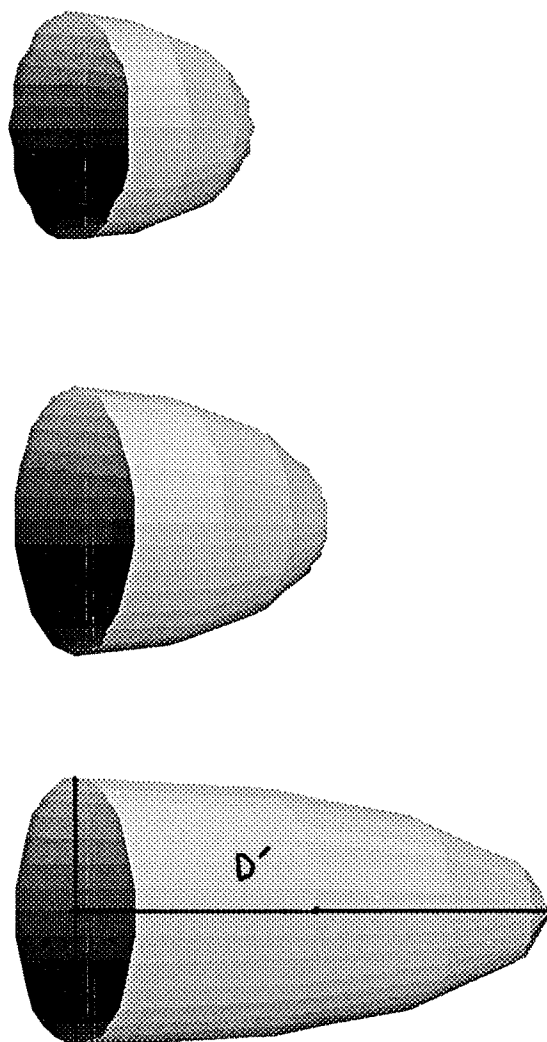
GENERAL METHOD

Unless otherwise indicated, a variation of a method used by Loomis and Eby (1988, Experiment 4) was also used in the present experiments. A one-parameter series of half-ellipsoids, varying only in the distance from base to apex, was developed (see Figure 1). Because many models of the human perception of SFM recover 3-D structure (e.g., Grzywacz & Hildreth, 1987; Landy, 1987; Ullman, 1984), Loomis and Eby (1990) have argued that dependent measures that capture local curvature or surface orientation are the best measures for making contact with theorizing. However, such methods are time consuming and difficult for observers. In the present study, a faster and less difficult means of judgment was chosen. The subject's task was to report the length, from base to apex, of the perceived half-ellipsoid (*D'* in Figure 1), under the assumption that this scalar measure captures the essential change in perceived structure produced by our manipulations. Despite the subjectivity of the measure, Loomis and Eby (1988, 1989, 1990) have shown that such depth estimates correlate highly with interesting measures of image motion (such as shearing motion), which suggests that scalar depth estimates are useful for theorizing about the human recovery of SFM.

Stimuli and Apparatus

The surfaces of the half-ellipsoids were defined by some set (usually 128 or 256) of randomly positioned luminous points, with the restriction that one of these points be positioned directly on the apex and four points evenly positioned about the base. Display point size was .4 mm (1.4'). So that subjects could not use unique features produced by the random placement of the points in the different objects as a basis for object identification, the half-ellipsoids were all created in the following way. Between objects, the *x* and *y* coordinates defining these points were identical; however, the *z* coordinate varied, depending on the simulated depth in the object. The SFM image sequences were created on an IBM PS/2 Model 80 microcomputer with software written for the 640 × 480 resolution mode of the IBM video graphics array (see Loomis & Eby, 1987b,

Three Examples of Half-Ellipsoids



3/4 View

Figure 1. Three examples of objects (half-ellipsoids) with shapes similar to those of the objects used in this study. For all objects, the base diameter was equal; the objects differed only in depth. D' shows the dimension judged by the subject. The objects were defined by points of light randomly positioned on their surfaces.

for a description of the software) and displayed on a Zenith flat screen RGB monitor (ZTM 1490). The video refresh rate was fixed at 60 Hz.

In most of the experiments, the objects were displayed on the CRT in the geometric configuration shown in Figure 2. The simulated object depicted in Figure 2 is the half-ellipsoid shown in Figure 1, oriented so that the base is parallel to the display screen with the apex recessed into the screen. The projected base of the

object was 7 cm in diameter (4° of visual angle). The projection center of the object base was always 5 cm from the screen center.

During each trial, the half-ellipsoid *translated curvilinearly* about the screen center,² and unless otherwise indicated, the *angular increment* was $1^\circ/\text{view}$. By analogy, the motion is similar to the orbital motion of the moon about the earth when viewed from above the earth's pole, with the moon as the object and the fixed point being the earth's center of mass.

This presentation method produced at least one source of depth information: velocity gradients based on motion parallax (the projected velocities varied with the distance to the projection point). The ability of the visual system to use this type of information is currently being investigated by several authors (e.g., Braunstein, Andersen, Rouse, & Tittle, 1986; Braunstein & Tittle, 1988; Loomis & Eby, 1990, 1991).

In the present study, we were interested in the buildup of 3-D structure as a function of the available motion information. This stimulus presentation method is advantageous for holding constant or eliminating the other monocular cues to depth that are often simultaneously present in a SFM display. As the object translated curvilinearly, it projected a nondeforming contour. Changes in the projected contour of an object in the absence of relative motion information have been shown to be a potentially useful source of information about 3-D object shape (e.g., Andersen & Cortese, 1989; Loomis & Eby, 1989; Pollick, 1989; Todd, 1985; Wallach & O'Connell, 1953). In addition, textural information, while possibly informative about object depth (e.g., Sperling, Landy, Doshier, & Perkins, 1989), is held constant while the object translates curvilinearly, allowing one to assess structural buildup without confounding the results with apparent depth produced by changing textural information. Curvilinear translation about the screen center, rather than some other type of translation, was also chosen because the simulated object center always remained at the same distance from the projection point used in generating the displays. Variations in this distance might have produced variations in perceived depth that were unrelated to the variables under investigation.

Range of motion and viewing duration were manipulated by varying α between trials. The display was viewed monocularly with the right eye and with head movement attenuated by a chinrest. The subject's eye was positioned on a line perpendicular to the center of the screen at a distance of 1 m. This viewing position was geometrically correct for the 1-m projection distance used in creating the 2-D imagery. A static view of the display appeared as a collection of white points contained within a circular area on a black background.

The experiments were under subject control and were run in a small darkened room. As depicted in Figure 3, the subject sat at a station with a chinrest. A computer keyboard was enclosed by a five-sided box that was open facing the subject. A small light source attached inside the box provided enough light for the subject to see the keyboard without the room's being visibly illuminated. The subject viewed the display screen over the box. A light source beneath the monitor illuminated a ruler that was calibrated in centimeters. The light source was connected to a foot pedal with which the subject could illuminate the ruler. The subject used the ruler to facilitate his or her judgment of object depth. The subject was kept in a moderate state of light adaptation by uniformly illuminating the display screen at a luminance of 65.1 cd/m^2 between trials.

Procedure

The experiments were divided into three or four sessions that were run with at least 2 h separating them. Prior to the first session, the subject was given instructions about the task and about how to perform the experiment. The subject was shown a diagram of the dimension of the half-ellipsoid that was to be judged (D' in Figure 1). The subject was told to respond to the apparent extent in depth rather than to what the subject "thought" was being shown. (The subject was judging the major axis of the half-ellipsoid; when the object base was oriented frontally, this coincided with extension in depth

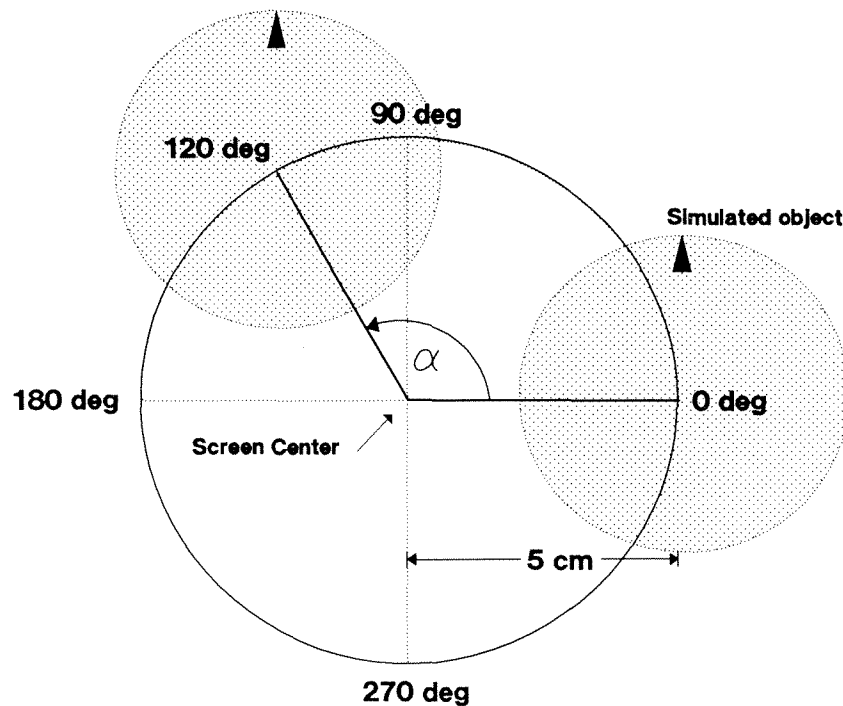


Figure 2. Schematic frontal depiction of the geometrical relationship of the factors involved in the display. The filled circles represent the simulated object in two different locations. The arrow in each filled circle shows that the object did not rotate as it translated. The larger circle indicates the path of curvilinear translation followed by the object and by the fixation target. Viewing duration and range of motion were varied by changing α between trials.

as seen by the observer.) Previous investigations in our laboratory have shown that these displays occasionally undergo depth reversals. Because this may have changed the perceived depth extent, subjects were instructed to attempt to maintain the depth ordering that they perceived at the beginning of the experimental sessions. When a depth reversal occurred during a trial, the subject was instructed to view the display until the depth reversed again before making a depth judgment. In postexperimental interviews, only about half of the subjects reported seeing spontaneous depth reversals. Of those who did report reversals, many indicated that the frequency of reversals tended to increase toward the end of the experimental sessions. No subjects reported being unable to "switch back" the depth ordering when a spontaneous reversal occurred. In the same interviews, the subjects were queried about possible strategies that they might have been using for their responses other than the requested strategy. All reported that the objects usually appeared to have a perceived depth, and all reported that they based their responses on this perception.

Each trial proceeded as follows. The subject pressed a key on the keyboard to begin. The adaptation display disappeared and a small (.4-mm) luminous fixation target appeared at either the 90° or the 270° position, as shown in Figure 2. This target translated curvilinearly around the screen center through an arc of 90° in a direction toward the 0° position in Figure 2. The subject was instructed to track this fixation target in order to initiate proper eye movements prior to stimulus presentation. When the target reached the 0° position, it was replaced with the stimulus, which continued to translate curvilinearly about the screen center in the same direction and with the same speed as the fixation target's. The stimulus moved through some range, α (depending on the particular trial), after which it disappeared. The subject was requested to judge the apparent depth in

the half-ellipsoid at its terminal position. Because in some of the trials the stimulus moved through a very short α , the viewing time may have been too short for subjects to form a judgment of the perceived depth. Therefore, after a 1-sec blanking period, the entire sequence was repeated.³ This cycle of stimulus display and blanking period continued until the subject was ready to make a judgment of depth. When ready, the subject terminated the display by pressing the appropriate key on the keyboard. The stimulus display was replaced

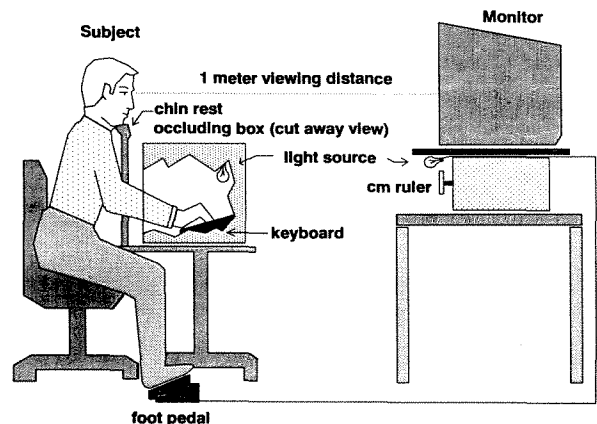


Figure 3. Schematic illustration of the apparatus used in all of the experiments, showing the location of the subject, monitor, and other components. (See the text for a description.)

by the adaptation display with a prompt for entering a response in centimeters. After the subject responded, the next trial began. Because several subjects participated in more than one experiment, no feedback about the hypotheses of the particular experiments was given until the end of the entire study.

EXPERIMENT 1 Speed of Motion

Experiment 1 was designed to investigate how the buildup of 3-D structure is affected by objects moving at different speeds. There were two main reasons for studying this variable. First, preliminary results (reported in Loomis & Eby, 1990) suggested that this variable affected the rate at which an object perceptually builds up. A second reason was to determine whether the perceptual buildup of structure was related to the range through which an object moved, to the display duration, or to both. By curvilinearly translating objects at several different speeds, we could study structural buildup as a function of display duration, independently of the range of motion.

Method

Subjects. Six graduate students from the University of California at Santa Barbara acted as observers and were paid for participating. As measured by a Keystone orthoscope, all observers had normal or corrected-to-normal visual acuity in the right eye (the eye used in the experiment). All observers were experienced in making psychophysical judgments of depth extent, but none were familiar with the hypotheses of the experiment.

Design. Three factors were investigated: simulated object depth (8, 16, 32, and 64 cm), range of motion (0° , 4° , 10° , 20° , 30° , 40° , 60° , and 120°), and speed of translation (30° , 60° , and $120^\circ/\text{sec}$). The perspective ratios (defined as in Braunstein, 1962) for the different depth objects were 1.08, 1.16, 1.32, and 1.64 for the shallowest to the deepest object, respectively. As depicted in Figure 4, translation speed (in degrees/second) was equal to the *update rate* multiplied by the *angular increment*. Translation speed was varied by independently manipulating these two factors to produce the three different speeds. The $120^\circ/\text{sec}$ speed was produced by curvilinearly translating an object through an increment of $2^\circ/\text{update}$ at an update rate of 60 Hz (note that the video refresh rate was fixed at 60 Hz). The slowest speed was produced by an angular increment of $1^\circ/\text{update}$ with a 30-Hz update rate. Since the fast and slow speeds were produced in different ways, to com-

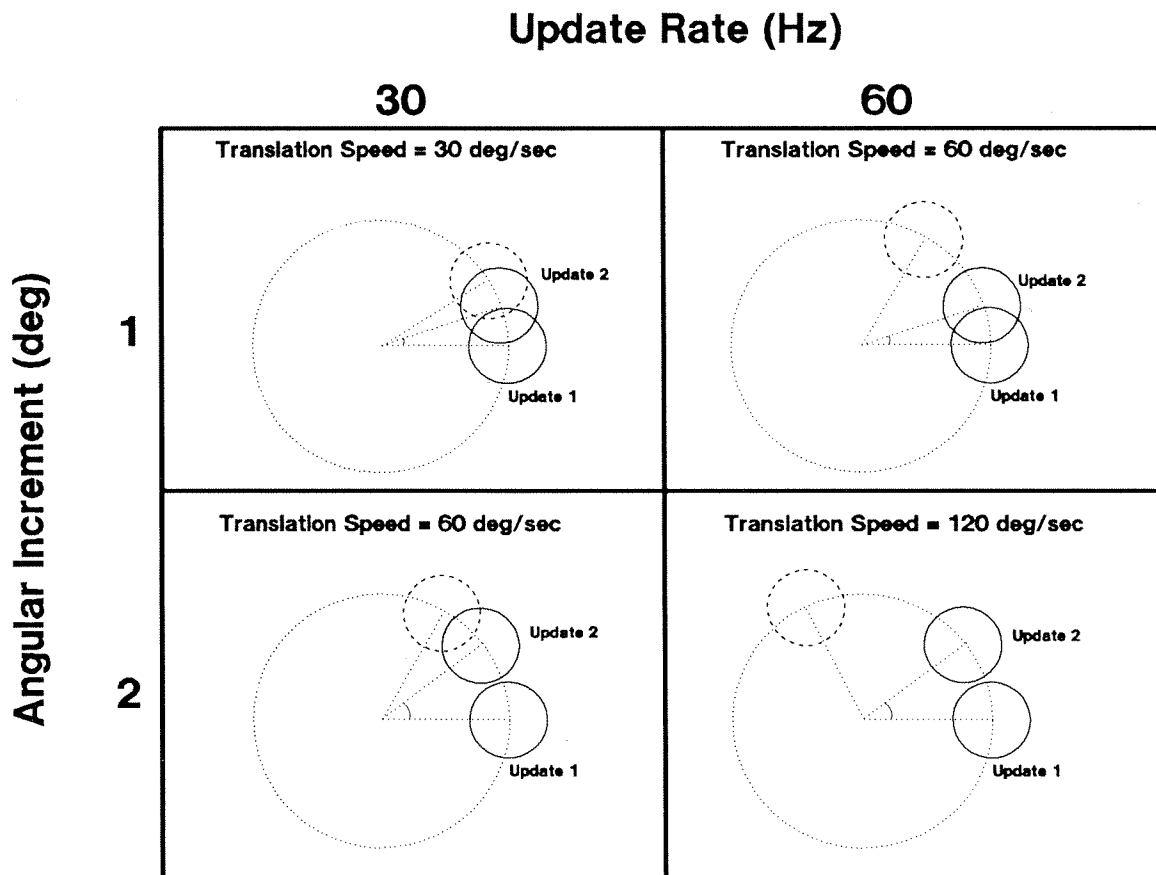


Figure 4. Schematic illustration of the relationship between the two factors manipulated to vary the curvilinear translation speed. Within each box, the large circle depicts the path of translation followed by the object. The small circles represent possible locations of the simulated object in various frames of the animation sequence. The small solid circles show that between updates the angular increment could be either 1° or 2° . (Note that the illustrated increments have been greatly exaggerated in this figure to make them noticeable.) The small dotted circle represents the location of the simulated object after 1 sec of display time; this location is the same in the upper right and lower left boxes, because these objects both translated at the same rate.

pare these conditions it was necessary to show that the two ways of varying the translation speed were functionally equivalent. Therefore, the 60°/sec curvilinear translation speed was produced in two ways: an angular increment of 1°/update with a 60-Hz update rate, and an angular increment of 2°/update with an update rate of 30 Hz. If no difference is observed in these two conditions, there is justification for comparing the different translation speeds.

The various combinations of these factors (4 object depths, 8 ranges of motion, and 4 translation speeds) resulted in 128 conditions in the experiment. The subject participated in three sessions; during each session, the subject judged all 128 trials in random order. Each session lasted approximately 40 min.

Results and Discussion

The average judgments of the 6 observers are shown in Figure 5. Each panel depicts a different curvilinear translation speed. Because the 60°/sec speed was produced in two different ways, two sets of data are shown in the middle panel. A three-factor (depth $\times$ range $\times$ speed) repeated measures analysis of variance (ANOVA) of the two 60°/sec speeds showed that there was no statistical difference between the two ways of curvilinearly translating an object at 60°/sec [$F(1,5) = .89, p = .39$], allowing us to compare across all speeds. Because no difference was observed in the 60°/sec speed conditions, in all subsequent analyses the data from these conditions have been averaged.

Several trends are evident in Figure 5. First, for all translation speeds, a main effect of simulated depth is evident. As is shown by the separation of the curves in each panel, the deeper the simulated object, the larger the judged extension in depth. A three-factor repeated measures ANOVA (depth $\times$ speed $\times$ range) showed this trend to be highly reliable [$F(3,15) = 28.51, p < .001$]. This result supports the findings of numerous other studies (e.g., Eby & Loomis, 1989; Loomis & Eby, 1988, 1989). Second, structural buildup was found in all conditions;

as the range of motion was increased, subjects judged the half-ellipsoid to be longer [$F(7,35) = 11.6, p < .001$]. This result is consistent with the findings of Hildreth et al. (1990) and Loomis and Eby (1988). Third, as the speed of translation was decreased, the average depth judgments decreased over all conditions. This main effect was marginally significant [$F(2,10) = 4.27, p < .05$]. Fourth, there was a reliable interaction between translation speed and range of motion [$F(14,70) = 5.26, p < .05$], indicating that between the translation speed conditions there was a slight difference in the shape of the curves that related reported depth and range of motion. Finally, as Hildreth et al. (1990) and Loomis and Eby (1988) found, across all translation speeds deeper simulated objects were judged as reaching asymptotic depth after about the same range of motion as that for shallow objects, but the levels of asymptotic depth differed. Thus, the curves relating judged depth and range of motion for the four simulated object depths had different shapes. This two-way interaction between simulated depth and range of motion was significant [$F(21,105) = 3.13, p < .001$]. All other interactions were nonsignificant.

As shown in Figure 5, the curves relating reports of depth and range of motion are approximately exponential in shape; this is the basic shape found in previous buildup experiments (Hildreth et al., 1990; Loomis & Eby, 1988). To compare the shapes of these curves, all curves were fit with the exponential function,

$$D' = E_{\max}[1 - (e^{-r/\sigma})], \quad (1)$$

where r is the range of motion in degrees, D' is the judged depth at range r , $E_{\max}$ is the asymptotic extension in depth, and σ is the space constant (the range of motion required for judged depth to come within $1 - 1/e$ of the maximum judgment of depth). The fitting was achieved by iteratively choosing the combination of $E_{\max}$ and σ that yielded the

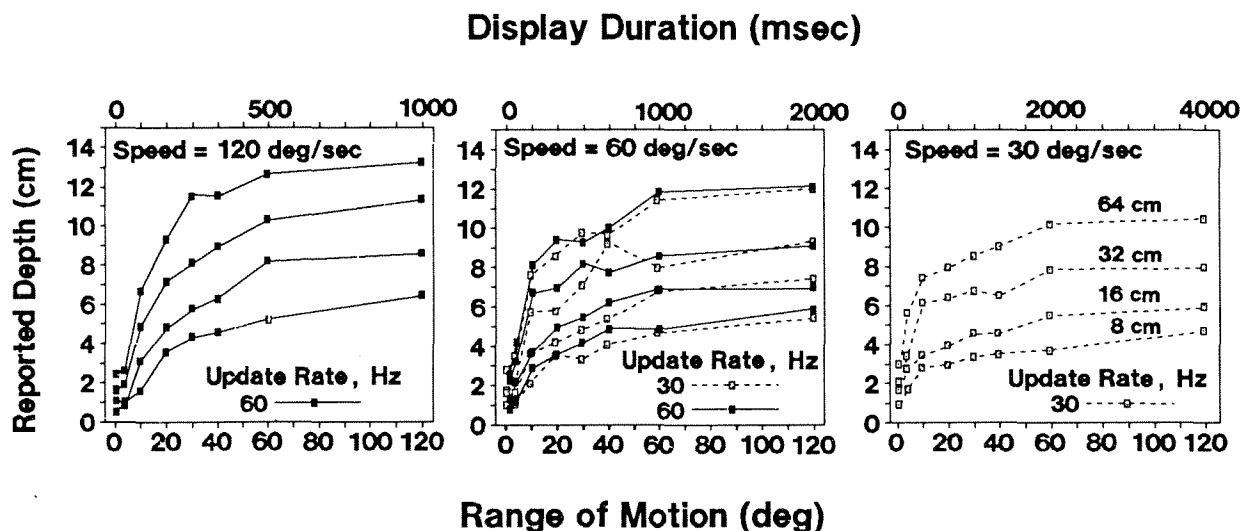


Figure 5. The results of Experiment 1, showing the average depth judgments of 6 observers as a function of simulated object depth, duration, range, and speed of translation.

Table 2
Calculated σ (in Degrees), τ (in Milliseconds), $E_{\max}$ (in Centimeters), and Root Mean Square (RMS) Error (in Centimeters) as a Function of Translation Speed and Simulated Object Depth (in Centimeters) for Experiment 1

Object Depth	Translation Speed											
	120°/sec				60°/sec				30°/sec			
	σ	τ	$E_{\max}$	RMS Error	σ	τ	$E_{\max}$	RMS Error	σ	τ	$E_{\max}$	RMS Error
8	35.1	291	6.6	.27	28.1	466	5.6	.22	18.7	626	4.2	.32
16	35.1	291	9.1	.47	26.6	446	7.3	.23	23.6	791	5.7	.26
32	27.5	231	11.4	.42	15.9	266	8.9	.53	10.4	351	7.4	.49
64	21.1	176	13.4	.67	17.3	291	11.7	.66	12.0	401	9.8	.55
Average	29.7	247			22.0	367			16.2	401		

Note—Both 60°/sec conditions were averaged before curve fitting.

lowest overall deviation from the empirical curve (calculated as root mean square, or RMS, error). (See the Appendix for a more detailed discussion of the curve-fitting procedure and the goodness of fit to the data.) The values derived with Equation 1 are shown in Table 2.

A review of Table 2 shows, as expected, that $E_{\max}$ increased with increases in simulated object depth. Interestingly, $E_{\max}$ also increased as translation speed increased. Without further research, we can only speculate about why this occurred. One possibility is that faster moving objects project a greater amount of relative optical motion information per unit of time. If we compare the space constant values, we find that σ did not vary systematically with simulated object depth in the 30°/sec and 60°/sec conditions, showing that the buildup of judged depth extent was approximately the same, regardless of simulated object depth. On the other hand, in the 120°/sec condition, σ decreased with increases in simulated depth, indicating that depth judgments built up at a faster rate when deeper objects were displayed. When we compare σ across translation speed conditions, we find that σ decreased with slower speeds, indicating that objects moving more slowly built up to maximum depth over a shorter range of motion than did fast moving objects. This finding suggests that not only spatial, but also temporal factors were involved in the buildup of structure.

Note that the results in Figure 5 are plotted as a function of both range of motion and display duration. If we average σ over the simulated object depths for each translation speed, we find average space constants of 29.7°, 22.0°, and 16.2° for the 120°/sec, 60°/sec, and 30°/sec conditions, respectively. If range of motion was the only factor related to structural buildup, we would expect these values to be nearly the same; this suggests that duration is also an important factor in the buildup of 3-D structure. When the data in Figure 5 are considered in terms of duration (note that each panel has a different time scale), it is clear that judgments of depth reach asymptotic depth levels at different durations. The curves in Figure 5 were fit with Equation 2,

$$D' = E_{\max}[1 - (e^{-t/\tau})], \quad (2)$$

where t is the duration in milliseconds and τ is the time constant. This curve fitting was identical to the procedure carried out with Equation 1, except that duration rather than range was used and a time constant, τ , rather than a space constant, σ , was fitted (see the Appendix). Averaging over simulated object depth, we find τ s of 247, 367, and 542 msec for the 120°/sec, 60°/sec, and 30°/sec conditions, respectively. If duration was the only factor related to structural buildup, we would expect little variation in these values. These facts suggest that *both* range of motion *and* stimulus duration are involved in the buildup of 3-D SFM. Without further experimentation, one cannot conclude definitively about the relative contributions of each factor; however, the fact that the best-fitting average σ values varied only slightly in comparison with the best fitting τ values suggests that the range through which an object moves is the more heavily weighted factor in the buildup of 3-D structure.

The remaining experiments were focused on other factors involved in the recovery of 3-D SFM, independently of the issue of spatial versus temporal effects. These factors were therefore covaried through the study of only one update rate. Because these factors were covaried, the results can be discussed in terms of either range of motion *or* duration; here the results are reported in terms of duration, so that different types of motions can be compared.

EXPERIMENT 2 Element Numerosity

The buildup of structure was investigated in Experiment 2 as a function of the number of texture elements defining the object. Past studies have shown that variations in the number of elements defining an object can affect the perceived rigidity of the object (Green, 1961; Todd, Akerstrom, Reichel, & Hayes, 1988), the ability of an observer to detect a change from an unstructured to a structured object (Husain, Treue, & Andersen, 1989), an observer's overall impression of three dimensionality (Braunstein, 1962), an observer's ability to identify object shape accurately (Doshier, Landy, & Sperling, 1989a), and an observer's impression of depth (Doshier, Landy,

& Sperling, 1989b). These studies suggested that one might observe a difference in the rate of structural buildup when element numerosity was varied.

Method

Subjects. Five subjects participated. Four were paid graduate student subjects recruited from the University of California at Santa Barbara and were unfamiliar with the hypotheses of the experiment. The 5th was the author. All subjects were experienced psychophysical observers who had normal or corrected-to-normal visual acuity in the right eye.

Design. Three variables were investigated: the simulated object depth (14, 20, 28, 40, and 56 cm), the number of points defining the simulated object (32, 64, 128, 256, and 512 points), and display duration (0, 67, 167, 334, 500, 668, 1,000, and 2,000 msec). These durations corresponded to α s of 0°, 4°, 10°, 20°, 30°, 40°, 60°, and 120°, respectively. The perspective ratios for the five objects were 1.14, 1.20, 1.28, 1.40, and 1.56 for the shallowest to the deepest object, respectively. In all objects, 4 points were equally spaced about the base of the half-ellipsoid, and a 5th point was positioned directly on the apex. The curvilinear translation speed was fixed at 60°/sec.

The combination of the variables yielded a total of 200 different conditions in the experiment. Each subject participated in three sessions, during which all 200 trials were judged in random order. The sessions were conducted on separate days, and each session lasted approximately 1.25 h.

Results

The average depth judgments of the 5 observers as a function of the display duration are shown in Figure 6.

Each panel depicts a different simulated object depth. As found in Experiment 1, subjects' judgments of apparent object depth increased with increasing simulated object depth. A three-factor (depth $\times$ duration $\times$ number of points) repeated measures ANOVA showed this main effect to be highly reliable [$F(4,800) = 65.23, p < .01$]. Also as in Experiment 1, judgments of depth increased with increases in duration up to about 1,000 msec [$F(7,800) = 120.69, p < .01$]. However, there was a significant tendency for shallower objects to continue building up after 1,000 msec [$F(28,800) = 2.65, p < .01$]. More interestingly, judgments of object depth were unaffected by the number of points defining the object [$F(4,800) = 3.89, p > .01$]. Because all other interactions were nonsignificant, one may conclude that for the range of point numerosities studied, varying the number of points that define an object does not affect the buildup of perceived 3-D structure.

The curves in Figure 6 were fit with Equation 2. τ , $E_{\max}$, and RMS error for each curve are shown in Table 3. As expected, $E_{\max}$ varied proportionally with simulated depth and did not vary systematically between point numerosities within each simulated depth condition. If we compare τ values, we find that, in general, τ decreased as simulated object depth increased, signifying that deeper simulated objects built up to asymptotic depth over a shorter duration than that for shallow objects; a similar trend was found in Experiment 1. In addition, it was found

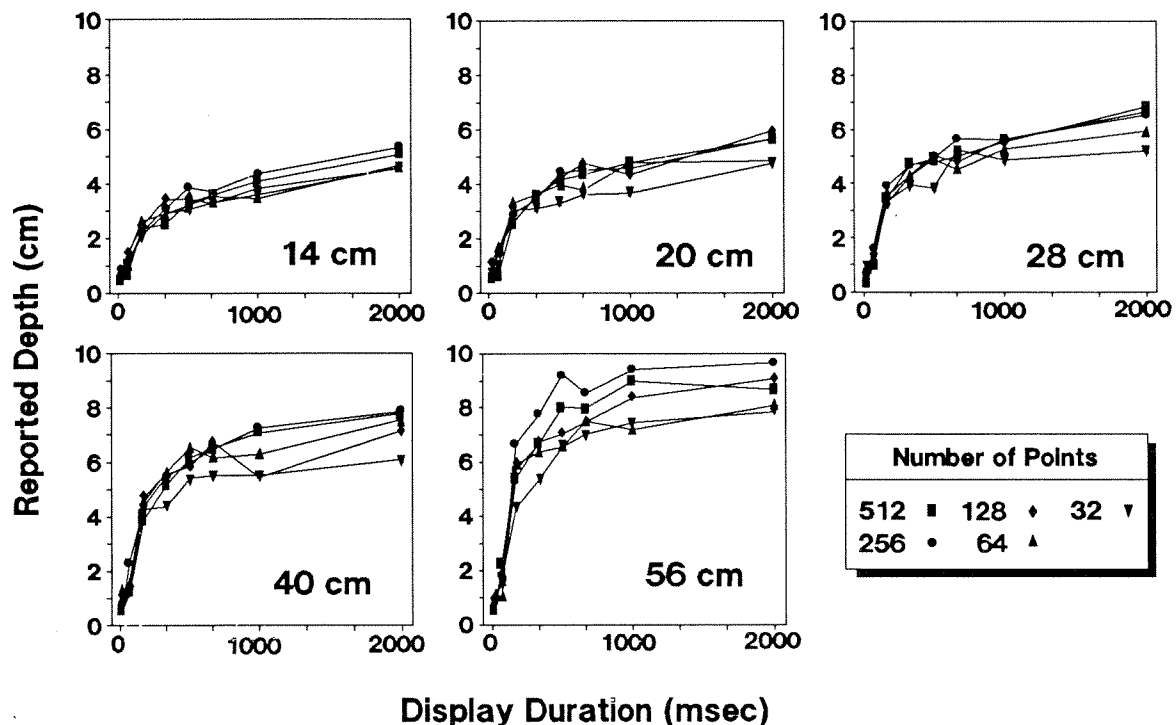


Figure 6. The results of Experiment 2, showing the average depth judgments of 5 observers as a function of the simulated object depth, number of points defining the object, and range of motion. The overlapping of the curves in each panel shows that there was little effect of point numerosity in the buildup of perceived structure.

Table 3
Calculated τ (in Milliseconds), $E_{\max}$ (in Centimeters), and Root Mean Square (RMS) Error (in Centimeters) as a Function of Simulated Object Depth (in Centimeters) and Element Numerosity for Experiment 2

Object Depth	Number of Points														
	512			256			128			64			32		
	τ	$E_{\max}$	RMS Error	τ	$E_{\max}$	RMS Error	τ	$E_{\max}$	RMS Error	τ	$E_{\max}$	RMS Error	τ	$E_{\max}$	RMS Error
14	541	5.0	.29	536	5.3	.23	262	4.1	.29	291	4.1	.35	486	4.4	.23
20	409	5.5	.32	346	5.4	.21	481	5.7	.32	271	4.7	.41	220	4.1	.40
28	292	6.5	.49	242	6.0	.35	362	6.3	.42	317	5.6	.36	287	5.2	.41
40	324	7.6	.33	254	7.4	.35	195	6.5	.65	291	7.1	.65	234	5.9	.48
56	209	8.7	.38	220	9.6	.82	257	8.6	.55	229	7.7	.85	314	7.8	.32
Average	355			320			311			280			308		

that τ did not vary systematically as a function of the number of points defining the object. The average τ values across each simulated depth condition for each point numerosity are shown in Table 3; these values had an average difference of only 32 msec, which was negligible, considering the fact that each view was updated every 16.7 msec. This finding supports the conclusion that varying the number of points that define an object does not affect the temporal buildup of perceived 3-D structure.

Discussion

This study showed that the buildup of perceived depth judgments in a SFM display is independent of the number of points that define the object, at least in the range studied. This fact is interesting for several reasons. First, the addition of points, from 32 to 512, does not slow the process of recovering SFM as might have been expected if the 3-D recovery process were based on a serial computation of image flow between all elements in the display. Second, even though element numerosity and spatial separation of points covaried in the study, it suggests that the process for recovering SFM can compute structure over a spatially extended range. In the 32-point conditions, the average separation between points was about 23.2 mm (1.3°), whereas in the 512-point conditions, this separation was only 5.7 mm ($.3^\circ$)—a fourfold difference. Alternatively, an average separation of 1.3° might represent a separation over which spatially local computation is sufficient, with smaller separations contributing little more. Research on the relationship between element numerosity and larger spatial separations in which the two do not covary should disentangle these possibilities. At any rate, for objects subtending a visual angle of 4° , there is no difference in the temporal rate of structural buildup or maximum judged depth level when the object is defined by either 32 or 512 points.

EXPERIMENT 3

Overlapping Versus Nonoverlapping Surfaces

Several studies have shown that the visual system can recover structure from the 2-D projections of 3-D objects that have a nonoverlapping surface (i.e., objects with a single smoothly varying surface, such as a plane or an

opaque spheroid) (e.g., Braunstein & Andersen, 1984; Eby, Loomis, & Solomon, 1989; Loomis & Eby, 1988; Todd, 1984) and from objects with overlapping transparent surfaces (such as a transparent spheroid) (e.g., Andersen, 1989; Braunstein, 1962; Braunstein & Andersen, 1984; Donner, Lappin, & Perfetto, 1984; Loomis & Eby, 1988, 1989; Mace & Shaw, 1974; Petersik, 1979). Experiment 3 was designed to investigate the buildup of structure as a function of whether the object has overlapping or nonoverlapping transparent surfaces. This variable was studied because the local projected flow fields are quite different in the two types of displays. In the case of an object with nonoverlapping surfaces, the optic flow field is locally smooth, whereas in the case of overlapping transparent surfaces, the local optic flow field is not smooth (i.e., points that are adjacent in the 2-D projection may be located on two surfaces separated in depth in the 3-D scene). This distinction is important, because mathematical analyses of optic flow have included assumptions about the smoothness of the flow field (e.g., Koenderink, 1986; Longuet-Higgins & Prazdny, 1980).

Andersen (1989) has investigated the ability of an observer to perceive 3-D structure with nonsmooth optic flow fields by simulating overlapping transparent planes separated in depth along the line of sight. He found that up to three surfaces separated in depth could be accurately detected and that judgments of depth between surfaces increased with increases in simulated depth separation. He also discovered that the sign of depth for two overlapping surfaces was accurately perceived. These results suggest that the visual system does not need a smooth optic flow field as input for recovering 3-D SFM. However, the display duration in all of Andersen's displays was 2 sec, a duration that typically yielded asymptotic depth judgments in the present study. It is possible that over the short term a smooth optic flow field is necessary, whereas over the longer term, processes not requiring locally smooth optic flow are used. If so, we would expect judgments of depth to increase more slowly for nonsmooth local optic fields than for smooth optic flow fields.

In this experiment, the paradigm from the first two experiments was used to investigate how perceived depth builds up as a function of whether the object has overlap-

ping (nonsmooth optic flow) or nonoverlapping transparent surfaces (smooth optic flow). In order to make the depth judgments comparable across objects, identically shaped objects were used. As shown in Figure 7, the nonoverlapping-surface objects were recessed half-ellipsoids with an open base; the objects with overlapping surfaces were also recessed half-ellipsoids, but the base was covered with points. With these specialized objects, the simulated extensions in depth were identical, allowing us to assess how depth builds up in the presence of overlapping or nonoverlapping transparent surfaces.

Method

Subjects. Five graduate students from the University of California at Santa Barbara participated for pay. Four of the subjects had participated in Experiment 1 or 2. All had experience in making psychophysical judgments of apparent depth in a SFM task. As measured by a Keystone orthoscope, all subjects had normal or corrected-to-normal visual acuity in the right eye. None of the subjects were familiar with the hypotheses of the experiment.

Design. Three factors were studied: simulated object depth (8, 16, 32, and 64 cm), the overlap in the surfaces defining the object (overlapping or nonoverlapping; see Figure 7), and the display duration (0, 67, 167, 334, 500, 668, 1,000, and 2,000 msec). These durations corresponded to α s of 0°, 4°, 10°, 20°, 30°, 40°, 60°, and 120°, respectively. The perspective ratios were 1.08, 1.16, 1.32, 1.64 for the shallowest to deepest object. The number of points per object was fixed at 256. For the nonoverlapping-surface objects, all 256 points were randomly distributed about the object surface with the restrictions discussed in the general method section. The overlapping-surface objects had 60 points randomly positioned on the base, and the remaining points (196) were randomly positioned over the rest of the object. The curvilinear translation speed was held constant at 60°/sec.

The various combinations of the three main factors produced a total of 64 different conditions in the experiment. Each subject par-

ticipated in four sessions; during each session, the subject judged all 64 conditions in random order. Each session lasted about 40 min.

Results and Discussion

The average judgments of the 5 observers are shown in Figure 8. Each panel depicts a different simulated object depth. As found in Experiments 1-2, increasing the simulated object depth and the display duration resulted in significant increases in reports of perceived depth. A 4 (simulated object depth) $\times$ 2 (degree of overlap) $\times$ 8 (duration) repeated measures ANOVA showed both trends to be highly reliable [$F(3,12) = 34.21$, $p < .001$, for depth; $F(7,28) = 9.49$, $p < .001$, for duration]. Additionally, there was a significant interaction between amount of surface overlap and simulated depth, reflecting the slight trend for subjects to report a greater difference in the apparent depths of the two kinds of objects as the simulated depth was increased [$F(3,12) = 4.64$, $p < .05$]. All other interactions were nonsignificant.

More interestingly, even though the graphs show that the overlapping-surface objects are judged to be shallower on the average than the nonoverlapping-surface objects in many cases, the difference is not statistically significant [$F(1,4) = 1.64$]. A comparison of individual subjects suggested that this observed difference in Figure 8 resulted mainly from the judgment of 1 observer, who judged nonoverlapping-surface objects to be as much as two times deeper than the overlapping-surface objects. This conclusion was supported by an ANOVA computed with subjects as a factor, which showed a statistically reliable interaction between the amount of overlap and subject variables [$F(4,12) = 13.3$, $p < .01$]. In addition, there was no significant main effect of amount of overlap

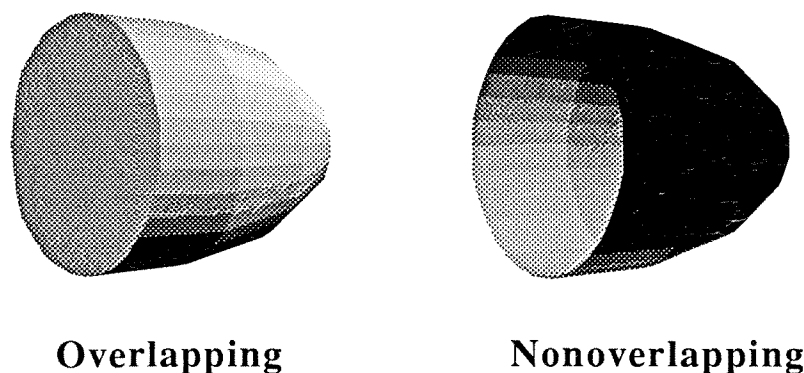


Figure 7. Depiction of the shapes used in Experiments 3 and 4. The nonoverlapping-surface half-ellipsoid is similar to the objects used in the other experiments in this study. The overlapping-surface object is also a half-ellipsoid; but, rather than having the base open, it was a surface defined by texture elements. The actual objects were transparent (rather than opaque as shown here) so that the impression was that of looking into the object through a transparent base. As in the other experiments, the objects were defined by points of light randomly distributed on the simulated surfaces.

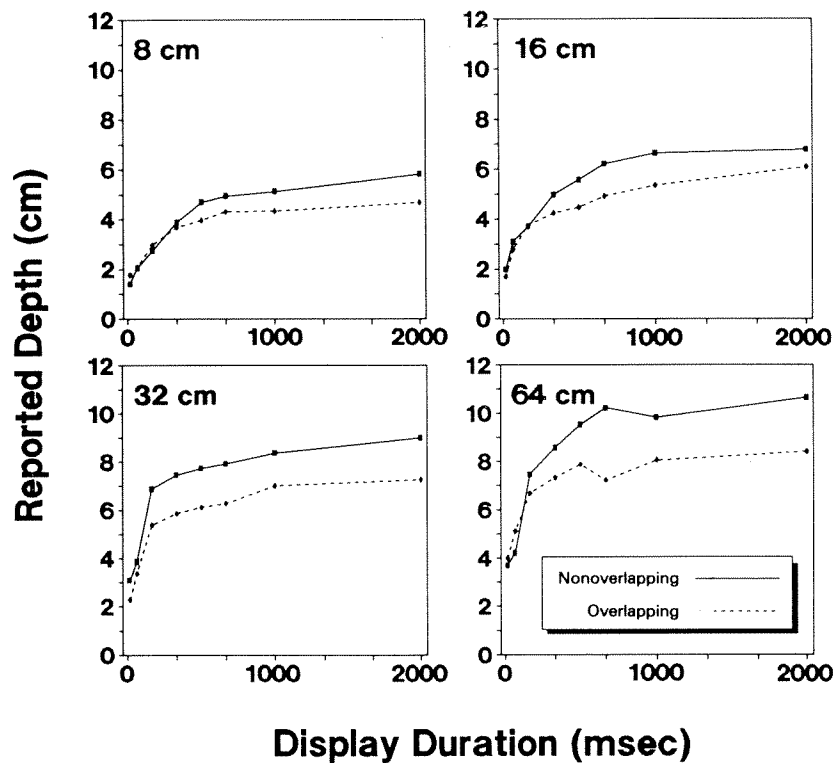


Figure 8. The results of Experiment 3, showing the average judgments of 5 observers as a function of the simulated object depth, the amount of surface overlap, and the range of motion. For all simulated object depths, there was a slight tendency for nonoverlapping objects to be judged as deeper, but this trend was not statistically significant.

[$F(1,3) = 6.1$]. This finding is consistent with the results of Andersen (1989). The observed separation of the curves in Figure 8 is likely to be the result of the judgments of 1 observer.

The eight curves in Figure 8 were fit with Equation 2. Table 4 shows τ , $E_{\max}$, and the goodness-of-fit measure (RMS error) for each curve. As found in the previous experiments, $E_{\max}$ increased as simulated object depth increased. Additionally, $E_{\max}$ was consistently larger in the nonoverlapping-surface conditions than in the overlapping-surface conditions; this was expected by looking at the

Table 4
Calculated τ (in Milliseconds), $E_{\max}$ (in Centimeters), and Root Mean Square (RMS) Error (in Centimeters) as a Function of Amount of Surface Overlap and Simulated Object Depth (in Centimeters) for Experiment 3

Object Depth	Surface Overlap					
	Nonoverlapping			Overlapping		
	τ	$E_{\max}$	RMS Error	τ	$E_{\max}$	RMS Error
8	403	5.75	.13	327	4.63	.10
16	351	6.87	.12	334	5.72	.30
32	205	8.53	.45	207	6.90	.32
64	259	10.50	.46	177	8.02	.30
Average	305			261		

graphs in Figure 8. In Experiment 2, τ was found to decrease with increases in simulated object depth. Overall, a similar trend was observed here, except in the case of the 64-cm simulated object. A comparison of τ between nonoverlapping- and overlapping-surface objects shows that in three of the four simulated depth conditions, τ was greater for the nonoverlapping-surface objects. Since there was no significant main effect for the overlap variable, this result was most likely produced by the observed subject $\times$ condition interaction, indicating only that at least 1 subject had a tendency to judge overlapping transparent surface objects as building up to maximum depth over a longer duration than that for nonoverlapping-surface objects. However, these differences consist only of a few degrees, and further research is needed to determine the effect of nonsmooth local optic flow on the buildup of perceived 3-D structure.

EXPERIMENT 4

The Effect of Rotation and Degree of Surface Overlap

In this experiment, structural buildup was investigated as a function of *rotation* about a vertical axis rather than curvilinear translation—as had been the case in the previous three experiments. There were two primary reasons for studying this factor. First, in all of Hildreth et al.'s

(1990) experiments, rotating displays were used. In the present experiment, buildup was investigated as a function of rotation so that the present findings could be better related to Hildreth et al.'s work. A second reason for studying rotational motion involved a SFM display developed by Jack Loomis at the University of California at Santa Barbara. In this display, an object rotated about a vertical axis. Every 180° of rotation, the simulation would instantaneously switch between a disk and a pill-shaped object. The impression as one viewed this display was that the perceived depth in the object changed rapidly and stabilized when the objects were switched; this stabilization period seemed to be much shorter than the .5 sec (60° range of translation) or so that we observed with curvilinearly translating objects in Experiments 1–3. Because this object was rotating rather than translating, it is possible that the buildup of structure is much faster when objects rotate. Additionally, because Experiment 3 showed a null effect of surface overlap but the graphs indicated a slight tendency for nonoverlapping-transparent-surface and overlapping-transparent-surface objects to perceptually build up at different rates, this factor was again varied.

Method

Subjects. Six graduate students from the University of California at Santa Barbara participated for pay. All had participated in at least one of the previous experiments. None were familiar with the hypotheses of the experiment.

Stimuli and Apparatus. The objects were created in the same way as in Experiment 3, except that they had a base radius of 5 cm. Since the objects were simulated as rotating about a vertical axis rather than translating, they were centered on the display screen. The objects rotated through various angles at an angular velocity of 60°/sec. The rotary increment between views of an object was 1°. As in the previous experiments, the displays were viewed from a distance of 1 m.

One difficulty in studying the recovery of 3-D structure from rotating nonsymmetrical objects is that as the object rotates, it projects a constantly deforming contour. As noted previously, changing contour information in isolation is sufficient for the recovery of shape (e.g., Andersen & Cortese, 1989; Loomis & Eby, 1989; Miles, 1931; Pollick, 1989; Todd, 1985; Wallach & O'Connell, 1953). This cue was therefore reduced in two ways. First, the range through which the objects rotated was limited, thereby minimizing the change in the projected contour. Second, the display was masked so that only a circular portion (radius = 4.5 cm) of the display was visible. The circular aperture was positioned so that it just occluded the circular border of the object. With these two manipulations, the projected contour of the objects was circular—except for the contour of the flattest object used in the experiment (a disk) when that object rotated through its greatest angular range (64°).

Design. Three factors were studied in this experiment: simulated depth of the object (0, 5, 10, and 15 cm), the amount of overlap in the surfaces defining the objects (overlapping or nonoverlapping; see Figure 8), and the display duration (0, 33, 67, 100, 134, 267, 534, and 1,069 msec). These durations corresponded to rotations about a vertical axis of 0°, 2°, 4°, 6°, 8°, 16°, 32°, and 64°, respectively. The perspective ratios were 1.00, 1.05, 1.10, and 1.15 for the shallowest to the deepest simulated objects. As shown in Figure 9, the rotary motion was evenly centered about the orientation in which the object base was parallel to the display screen.

The display procedure was similar to that employed in the previous studies. The display started with the object oriented at its maximum angular extent (Position 1 in Figure 9). It then rotated through

the appropriate angular range for that trial to its other maximum extent (Position 2 in Figure 9). The display was then blanked for 1 sec. This cycle of stimulus display and blanking was continually repeated until the subject responded.

The various combinations of these factors yielded 64 different conditions for the experiment. The subject participated in four separate sessions, during which he or she judged all 64 conditions in random order. Each session lasted about 40 min.

Results

The average results of the 6 observers are shown in Figure 10. Each panel depicts a different simulated object depth. As in all of the previous studies, there were systematic effects of simulated object depth and display duration, with reports of depth increasing with increases in either factor. A 4 (depth) $\times$ 8 (duration) $\times$ 2 (amount of overlap) repeated measures ANOVA showed both effects to be highly reliable [$F(3,15) = 48.82, p < .001$, for depth; $F(7,35) = 42.22, p < .001$, for duration]. Additionally, a review of the panels in Figure 10 shows that as simulated depth increased, the structure required a longer duration to build up. This interaction was significant [$F(21,105) = 22.81, p < .001$]. In support of the null effect of surface overlap found in Experiment 3, no main effect of surface overlap was observed [$F(1,5) = .44$]. All other comparisons were nonsignificant.

The eight curves in Figure 10 were fit with Equation 2. τ , $E_{\max}$, and goodness-of-fit values (RMS error) for each curve are shown in Table 5. As found in all of the previous experiments, $E_{\max}$ increased with larger simulated object depths. In support of the findings in Experiment 3, $E_{\max}$ was consistently larger for the objects with nonoverlapping surfaces. A comparison of the τ values shows that as deeper objects are simulated, the derived τ values increase slightly, indicating that deep objects built up to maximum depth over a greater duration than that for shall-

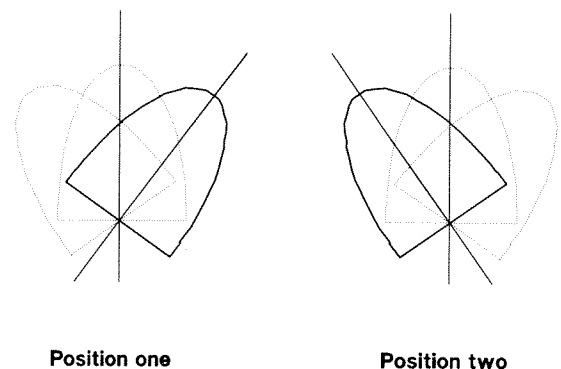


Figure 9. Schematic illustration showing the positioning of the half-ellipsoids in Experiment 4. This top view shows the two extreme orientations for an example half-ellipsoid. During each trial, the object began at Position 1 and then rotated with a constant velocity about a vertical axis until it reached Position 2. In this diagram, the rotation axis is orthogonal to the page and coincident with the intersection of the two dotted lines. These two positions were always symmetrical with the line of sight.

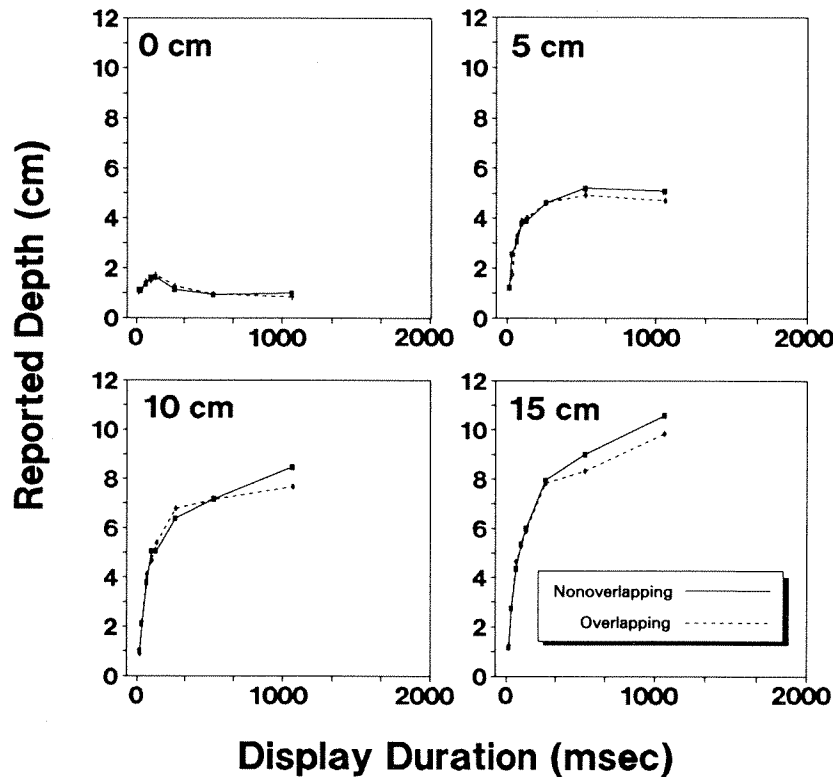


Figure 10. The results of Experiment 4, showing the average depth judgments of 6 observers as a function of the simulated object depth, the amount of surface overlap, and the rotation range. The similarity of the lines in each panel shows that there was no effect of surface overlap replicating the results in Experiment 3. In the 5- and 10-cm simulated object depths, it was found that judgments of object depth built up to maximum in only about 16° of rotation, a duration of about 270 msec. This buildup is much faster than that observed for curvilinearly translating objects.

low objects. This is surprising, because it is opposite to the trend that was found in Experiments 1–3 with curvilinearly translating objects. Comparing τ across the surface overlap variable shows that τ was slightly, but consistently, higher for the nonoverlapping objects, signifying a more rapid buildup in the nonoverlapping-surface objects. This finding is consistent with the small trend found in Experiment 3, suggesting that the small but reliable differences in the τ values between the amount of over-

Table 5
Calculated τ (in Milliseconds), $E_{\max}$ (in Centimeters), and Root Mean Square (RMS) Error (in Centimeters) as a Function of Amount of Surface Overlap and Simulated Object Depth (in Centimeters) for Experiment 4

Object Depth	Surface Overlap					
	Nonoverlapping			Overlapping		
	τ	$E_{\max}$	RMS Error	τ	$E_{\max}$	RMS Error
0	23	1.27	.25	20	1.28	.10
5	100	5.07	.14	90	4.85	.24
10	142	7.86	.40	115	7.47	.20
15	169	10.05	.35	145	9.24	.38
Average	109			93		

lap conditions *may* reflect some basic property of the process involved in the recovery of 3-D SFM. However, further research is needed to test this possibility.

When the τ values in Experiment 3 (Table 4) are compared with the τ values in Experiment 4 (Table 5), we find that when objects are rotating, judgments of depth build up to maximum over a much shorter duration. Collapsing across the surface overlap variable for the 16-cm object in Experiment 3 and the 15-cm object in Experiment 4 (objects that were closely matched in perspective), we find average τ s of 343 and 157 msec, respectively. This difference is more than a factor of two, showing that judgments of perceived object depth for rotating objects build up to a maximum in half the time required for curvilinearly translating objects.

Discussion

When the results of Experiment 4 are compared with the results of Experiment 3, we find that structure builds up much more rapidly when objects are simulated as rotating in depth rather than curvilinearly translating. This is the likely explanation for the fast structural buildup of the rotating object observed in the object-switching display described earlier.

Why do judgments of rotating objects build up faster than judgments of translating objects? Without further research, this question cannot be answered definitively. One possibility, however, is that different processes for recovering 3-D depth are used when objects are rotating than when they are translating (e.g., Braunstein, 1986; Braunstein & Andersen, 1984; Braunstein et al., 1986; Braunstein & Tittle, 1988). Support for this idea is given by the facts that (1) the trends for the time constants between the different simulated object depths were opposite in direction for rotation and for curvilinear translation and (2) a comparison of the $E_{\max}$ values for objects of similar simulated depths showed that the judgments of depth extension leveled off at a much higher value for rotating objects than for objects translating along a curvilinear path (see Table 4, 16-cm object, and Table 3, 15-cm object).

GENERAL DISCUSSION

In Experiment 1, it was found that the buildup of 3-D structure was dependent on the speed at which an object curvilinearly translated as well as the range through which the object translated. Experiment 2 showed that the buildup of depth judgments was unaffected by the number of points defining the object, including the maximum apparent depth within each simulated object size condition. This null effect of point numerosity on maximum judgments of apparent depth is consistent with previous findings (Eby & Loomis, 1989; Doshier et al., 1989b). Experiments 3 and 4 showed that structural buildup is essentially the same regardless of the smoothness of the optic flow field; however, a comparison of the best-fitting τ values indicated a tendency for objects that produced a smooth local optic flow (nonoverlapping-surface objects) to build up more slowly than objects that produced nonsmooth local optic flow (overlapping-transparent-surface objects). In Experiment 4, the effect of structural buildup was studied as a function of rotary motion rather than curvilinear translation. It was found that rotary motion produced a much more rapid increase in judgments of apparent depth than did curvilinear translation; the average τ was 127 msec, less than half the average τ in the fastest object translation condition in Experiment 1.

In all of the experiments, judgments of apparent depth reached asymptotic values at a level that was less than the simulated depth in the objects; this trend was strongest for the objects that were simulated as translating curvilinearly about the screen center. Such underestimations of apparent depth are common (see, e.g., Andersen, 1989; Braunstein & Tittle, 1988; Eby & Loomis, 1989; Loomis & Eby, 1987a, 1988, 1989, 1990).

There are several possibilities for why this occurred. One is that the subject is underestimating the absolute distance to the display monitor (Andersen, 1989; Ono, Rivest, & Ono, 1986), resulting in a foreshortening of the perceived depth from relative optical motion. Since no data about the absolute distance to the display was col-

lected, we cannot rule out this possibility. However, it seems unlikely that this could account for *all* of the foreshortening of perceived depth extent reported here. For example, in Experiment 3, the deepest simulated object was 64 cm in depth, yet judgments of depth averaged only about 10 cm; this is an underestimation of about 83%. Such underestimations of depth would likely require the subject to perceive the absolute distance of the display to be at least two thirds closer than its physical location. This is unlikely, and it is not supported by the subjective impressions of the author.

As suggested by Braunstein and Tittle (1988), another possibility is that other sources of information signaling flatness work against the relative optical motion information signaling depth. In the present displays, the size and luminance of all texture elements were equal, indicating a flat object; accommodation would have provided information that the display was flat; and the equidistance tendency (Gogel, 1965) might have provided a flatness signal. Although the effectiveness of this type of information to signal flatness in SFM displays has not been systematically investigated, it is likely that they reduced the subjective impressions of depth in the present displays.

However, the characteristics of the texture elements, accommodation, the equidistance tendency, and presumably the absolute distance to the display were constant throughout this study. If these sources of information about flatness do not vary in effectiveness between different kinds of motions, and if the underestimation of depth was entirely the result of these factors, one would expect to find objects of similar simulated depths to be judged as appearing to be roughly equal in depth extent when the objects undergo different motions. This was not the case in the present study; rotating objects built up to a greater depth than did curvilinearly translating objects (compare Figures 8 and 10). This finding is consistent with the results of Loomis and Eby (1990), who have shown that perceived depth judgments of the same simulated object undergoing a wide variety of motions vary systematically with the simulated motions, even when the flatness information discussed previously is constant throughout the experiment. These results suggest that (1) the effectiveness of flatness information to signal flatness may vary as a function of the simulated 3-D motion; (2) the depth information produced by the different simulated 3-D motions may vary in effectiveness; or (3) a combination of both possibility 1 and possibility 2 contributes to the foreshortening of depth judgments.

The curves relating reported depth to display duration in each experiment were fit with exponential functions (Equations 1 and 2). Two consistent trends were observed between experiments. First, $E_{\max}$ increased with larger simulated object depths, as expected. Second, in Experiments 1, 2, and 3, there was a tendency for the derived τ values to vary inversely with simulated object depth. This fact suggests that deep simulated objects built up to maximum depth over a shorter duration (and range of motion) than did shallow objects when objects were simu-

lated as curvilinearly translating. Because the 2-D projections of deep objects in motion produce a greater amount of relative motion between image points than do shallow objects undergoing the same motion, the trend for simulated object depth and τ may suggest that the output of the mechanism for recovering 3-D SFM is closely linked with the magnitude of relative motion in a display.⁴ If so, we would expect the objects defined by overlapping transparent surfaces (Experiments 3 and 4) to be best fit by τ values that are smaller than the best-fitting τ s for non-overlapping-surface objects. (The reason for this is that, for overlapping-transparent-surface objects, texture elements on one surface move relative to the elements on the transparent surface in front; this is relative motion that would not be present without the second surface.) Indeed, both experiments on the amount of overlap of transparent surfaces showed slightly higher τ values for nonoverlapping-surface objects; however, these differences were quite small.

The nature of the relationship between τ and relative motion, however, is still unclear. A simple measurement of relative motion in a display (e.g., Loomis & Eby, 1989, 1990) cannot explain why opposite effects for τ and simulated object depth were found when objects rotated in Experiment 4. In this experiment, τ increased with increases in simulated object depth, a trend opposite to that observed in Experiments 1, 2, and 3. This finding supports the idea that optic flow fields produced by rotation may be processed differently than optic flow fields produced by translations (see, e.g., Braunstein, 1986; Braunstein & Andersen, 1984), or, at the very least, analyzed using information other than relative motion in the interpretation of the optic flow. Moreover, in Experiment 1, τ decreased with decreasing speeds of translation. The fact that this decrease in τ occurred even when the average relative motion between points per degree of curvilinear translation was the same for all three translation speeds suggests that the mechanism for recovering 3-D SFM may respond to the time rate of relative motion rather than simply relative motion per distinct view.

In summary, the present results agree with and extend the results of previous investigations into the buildup of 3-D structure (Hildreth et al., 1990; Loomis & Eby, 1988). The results show that Wallach and O'Connell's (1953) assertion that tridimensionality is *immediately* perceived in a KDE display is correct. However, judgments of amount of depth continue to increase up to about .5 sec. In addition, the rate of this buildup is dependent on the type of object motion, the duration, and the range through which the object moves.

REFERENCES

- ANDERSEN, G. J. (1989). Perception of three-dimensional structure from optic flow without locally smooth velocity. *Journal of Experimental Psychology: Human Perception & Performance*, **15**, 363-371.
- ANDERSEN, G. J., & CORTESI, J. M. (1989). 2-D contour perception resulting from kinetic occlusion. *Perception & Psychophysics*, **46**, 49-55.
- BRAUNSTEIN, M. L. (1962). Depth perception in rotating dot patterns: Effects of numerosity and perspective. *Journal of Experimental Psychology*, **64**, 415-420.
- BRAUNSTEIN, M. L. (1986). Dynamic stereo displays for research on the recovery of three-dimensional structure. *Behavior Research Methods, Instruments, & Computers*, **18**, 522-530.
- BRAUNSTEIN, M. L., & ANDERSEN, G. J. (1984). Shape and depth perception from parallel projections of three dimensional motion. *Journal of Experimental Psychology: Human Perception & Performance*, **10**, 749-760.
- BRAUNSTEIN, M. L., ANDERSEN, G. J., ROUSE, M. W., & TITTLE, J. S. (1986). Recovering viewer-centered depth from disparity, occlusion, and velocity gradients. *Perception & Psychophysics*, **40**, 216-224.
- BRAUNSTEIN, M. L., & TITTLE, J. S. (1988). The observer-relative velocity field as the basis for effective motion parallax. *Journal of Experimental Psychology: Human Perception & Performance*, **14**, 582-590.
- DONNER, J., LAPPIN, J. S., & PERFETTO, G. (1984). Detection of three-dimensional structure in moving optical patterns. *Journal of Experimental Psychology: Human Perception & Performance*, **10**, 1-11.
- DOSHER, B. A., LANDY, M. S., & SPERLING, G. (1989a). Kinetic depth effect and optic flow—I. 3D shape from Fourier motion. *Vision Research*, **29**, 1789-1813.
- DOSHER, B. A., LANDY, M. S., & SPERLING, G. (1989b). Ratings of kinetic depth in multidot displays. *Journal of Experimental Psychology: Human Perception & Performance*, **15**, 816-825.
- EBY, D. W., & LOOMIS, J. M. (1989). The minimal effect of occluding stripes on the perception of structure from motion. *Investigative Ophthalmology & Visual Science*, **30**(Suppl.), 251.
- EBY, D. W., LOOMIS, J. M., & SOLOMON, E. M. (1989). Perceptual linkage of multiple objects rotating in depth. *Perception*, **18**, 427-444.
- GOGEL, W. C. (1965). The equidistance tendency and its consequences. *Psychological Bulletin*, **64**, 153-163.
- GREEN, B. F. (1961). Figure coherence in the kinetic depth effect. *Journal of Experimental Psychology*, **62**, 272-282.
- GRZYWACZ, N. M., & HILDRETH, E. C. (1987). The incremental rigidity scheme for recovering structure from motion: Position-based vs. velocity-based formulations. *Journal of the Optical Society of America A*, **4**, 503-518.
- HILDRETH, E. C., GRZYWACZ, N. M., ADELSON, E. H., & INADA, V. K. (1990). The perceptual buildup of three-dimensional structure from motion. *Perception & Psychophysics*, **48**, 19-36.
- HILDRETH, E. C., & KOCH, C. (1987). The analysis of visual motion: From computational theory to neuronal mechanisms. *Annual Review of Neuroscience*, **10**, 477-533.
- HUSAIN, M., TREUE, S., & ANDERSEN, R. A. (1989). Surface interpolation in three-dimensional structure-from-motion perception. *Neural Computation*, **1**, 324-333.
- INADA, V. K., HILDRETH, E. C., GRZYWACZ, N. M., & ADELSON, E. H. (1986). The perceptual buildup of three-dimensional structure from motion. *Investigative Ophthalmology & Visual Science*, **26**(Suppl.), 142.
- KOENDERINK, J. J. (1986). Optic flow. *Vision Research*, **26**(Suppl.), 161-180.
- LANDY, M. S. (1987). A parallel model of the kinetic depth effect using local computations. *Journal of the Optical Society of America A*, **4**, 864-876.
- LANDY, M. S., DOSHER, B. A., SPERLING, G., & PERKINS, M. E. (1988). *The kinetic depth effect and optic flow: II. Fourier and non-Fourier motion* (Mathematical Studies in Perception & Cognition, Report No. 88-4). New York: New York University.
- LONGUET-HIGGINS, H. C., & PRAZDNY, K. (1980). The interpretation of a moving retinal image. *Proceedings of the Royal Society: Series B*, **208**, 385-397.
- LOOMIS, J. M., & EBY, D. W. (1987a). Perceiving 3-D structure from motion: The importance of axis of rotation. *Investigative Ophthalmology & Visual Science*, **28**(Suppl.), 234.
- LOOMIS, J. M., & EBY, D. W. (1987b). High-speed 2-D and 3-D ani-

- mation on the IBM PC/XT/AT. *Behavior Research Methods, Instruments, & Computers*, **19**, 10-18.
- 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* (pp. 383-391). Washington DC: IEEE.
- LOOMIS, J. M., & EBY, D. W. (1989). Relative motion parallax and the perception of structure from motion. In *Proceedings of the IEEE Workshop on Visual Motion* (pp. 204-211). Washington, DC: IEEE.
- LOOMIS, J. M., & EBY, D. W. (1990). The dependence of perceived shape on object motion. *Investigative Ophthalmology & Visual Science*, **31**(Suppl.), 172.
- LOOMIS, J. M., & EBY, D. W. (1991). Velocity gradients and perceived slant. *Investigative Ophthalmology & Visual Science*, **32**(Suppl.), 958.
- MACE, W. M., & SHAW, R. (1974). Simple kinetic information for transparent depth. *Perception & Psychophysics*, **15**, 201-209.
- MILES, W. R. (1931). Movement interpretations of the silhouette of a revolving fan. *American Journal of Psychology*, **43**, 392-405.
- ONO, H., RIVEST, J., & ONO, H. (1986). Depth perception as a function of motion parallax and absolute-distance information. *Journal of Experimental Psychology: Human Perception & Performance*, **12**, 331-337.
- PETERSIK, J. T. (1979). Three-dimensional object constancy: Coherence of a simulated rotating sphere in noise. *Perception & Psychophysics*, **25**, 328-335.
- POLLOCK, F. E. (1989). Shape perception from dynamic occluding contours. *Investigative Ophthalmology & Visual Science*, **30**(Suppl.), 264.
- SERLING, G., LANDY, M. S., DOSHER, B. A., & PERKINS, M. E. (1989). The kinetic depth effect and identification of shape. *Journal of Experimental Psychology: Human Perception & Performance*, **15**, 826-840.
- TODD, J. T. (1984). The perception of three-dimensional structure from rigid and nonrigid motion. *Perception & Psychophysics*, **36**, 97-103.
- TODD, J. T. (1985). The perception of structure from motion: Is projective correspondence of moving elements a necessary condition? *Journal of Experimental Psychology: Human Perception & Performance*, **11**, 689-710.
- TODD, J. T., AKERSTROM, R. A., REICHEL, F. D., & HAYES, W. (1988). Apparent rotation in three-dimensional space: Effects of temporal, spatial, and structural factors. *Perception & Psychophysics*, **43**, 179-188.
- TODD, J. T., & BRESSAN, P. (1990). The perception of 3-dimensional affine structure from minimal apparent motion sequences. *Perception & Psychophysics*, **48**, 419-430.
- ULLMAN, S. (1979). *The interpretation of visual motion*. Cambridge, MA: MIT Press.
- ULLMAN, S. (1984). Maximizing rigidity: The incremental recovery of 3-D structure from rigid and nonrigid motion. *Perception*, **13**, 255-274.
- WALLACH, H., & O'CONNELL, D. N. (1953). The kinetic depth effect. *Journal of Experimental Psychology*, **45**, 205-217.

NOTES

1. An earlier report of this research can be found in Inada, Hildreth, Grzywacz, and Adelson (1986).
2. In previous studies (Loomis & Eby, 1988, 1989), we have called this type of motion *revolution*.
3. From the results of pilot studies, it was determined that this blanking period was of sufficient duration to allow any recovered depth to collapse before the next stimulus display. The subject was allowed to view the stimulus for as many repetitions as was necessary to make an assessment of apparent depth extent; the importance of unlimited viewing time has been pointed out by Todd and Bressan (1990).
4. A similar idea has been pursued by Loomis and Eby (1989, 1990), who have shown that judgments of depth (after asymptotic depth levels were reached) tend to correlate highly with a global measure of relative motion parallax calculated on the display elements.

APPENDIX

In order to describe the shapes of the curves in this study, all of the curves were fit with exponential functions of the form:

$$D' = E_{\max}[1 - (e^{-r/\sigma})],$$

where D' is the judged perceived depth at some range, r is the range of motion in degrees, $E_{\max}$ is the asymptotic extension in depth, and σ is the space constant (in degrees) denoting the steepness of the curve. Written in terms of temporal variables, this equation is:

$$D' = E_{\max}[1 - (e^{-t/\tau})],$$

where t is the duration (in milliseconds) and τ is the fitted time constant (in milliseconds).

For each empirical curve, the best-fitting exponential function was determined by iteratively selecting the combination of $E_{\max}$ and σ or τ that minimized the RMS error from the empirical data. $E_{\max}$ was varied because in many of the curves a definite asymptotic depth value was not apparent. Over all of the experiments, the fits of the curves to the data were quite good; RMS error values ranged from .10 to .85 cm, with most values falling around .50 cm.

(Manuscript received November 5, 1990;
revision accepted for publication September 18, 1991.)