Vision Res. Vol. 31, No. 1, pp. 59–75, 1991
Printed in Great Britain.

0042-6989/91 $3.00 + 0.00

HUMAN PERCEPTION OF STRUCTURE FROM MOTION

STEFAN TREUE,[1] MASUD HUSAIN[2] and RICHARD A. ANDERSEN[1]*

[1]Massachusetts Institute of Technology, Department of Brain and Cognitive Sciences, E25-236, Cambridge, MA 02139, U.S.A. and [2]Wolfson College, University of Oxford, Linton Road, Oxford OX2 6UD, U.K.

(*Received* 2 *October* 1989; *in revised form* 18 *April* 1990)

Abstract—Novel dynamic random-dot displays representing a rotating cylinder or a noise-field were used to investigate the perception of structure from motion (SFM) in humans. The finite lifetimes of the points allowed the study of spatiotemporal characteristics with smoothly moving stimuli. In one set of experiments subjects had to detect the change from the unstructured motion to the appearance of the cylinder in a reaction time task. In another set of experiments subjects had to distinguish these two stimuli in a two-alternative forced-choice task. The two major findings were: (1) a relatively constant point lifetime threshold (50–85 msec) for perceiving structure from motion. This threshold is similar to the threshold for estimating velocity and suggests that velocity measurements are used to process SFM; (2) long reaction times for detecting structure (~ 1 sec). The build-up of performance with time and with increasing numbers of points reflects a process of temporal and spatial integration. We propose that this integration is achieved through the generation of a surface representation of the object. Information from single features on the object appears to be used to interpolate a surface between these local measurements allowing the system to improve perception over extended periods of time even though each feature is present only briefly. Selective masking of the stimulus produced characteristic impairments which suggest that both velocity measurements and surface interpolation are global processes.

Structure from motion Velocity Surface interpolation Computational algorithms

INTRODUCTION

It has long been appreciated that humans are capable of perceiving the three-dimensional shape of an object using motion cues (Miles, 1931). The shadow of a static, bent paper-clip projected onto a 2-D screen, for example, offers little insight into its 3-D shape. Yet, if the clip is rotated so that the shadows of its parts move relative to each other, its 3-D structure becomes immediately apparent (Wallach & O'Connell, 1953)—the so-called kinetic depth effect or "structure from motion" (SFM).

Geometrical considerations show that the recovery of 3-D structure from motion from 2-D images is not a trivial problem. Since there are an infinite number of 3-D interpretations of a given pattern of motion in a 2-D image, the problem is considered to be "ill-posed" (Poggio & Koch, 1985). In order to formulate a unique, one-to-one mapping between the 2-D and the 3-D interpretation constraints are required. Different constraints have been proposed to restrict the range of solutions and have led to the development of a number of computational theories which specify a unique 3-D interpretation for moving elements in a series of 2-D images.

Two general classes of theory have been proposed: those that use velocity information ("velocity algorithms" or "continuous algorithms") and those that employ position measurements ("position algorithms" or "discrete algorithms"). Since both types of theory claim to solve the SFM problem, the question arises as to which, if any, of the several proposed computational algorithms may be employed by the human visual system.

Although several investigators have reported observations regarding the perception of SFM, it has been difficult to use these findings to determine which of these proposed algorithms might be used by the human visual system. Three factors have contributed to this difficulty. First, many of the stimuli used contained non-motion cues, of which the most important is the presence of patterns in the displays which

*To whom reprint requests should be addressed.

change their 2-D shapes during rotation. Some stimuli used in previous studies of SFM perception contained nonmotion cues which allowed subjects to perform above chance in the absence of motion (Braunstein, Hoffman, Shapiro, Andersen & Bennett, 1987). The second factor which makes the interpretation of many previous studies difficult is the use of informal observations or subjective assessments of the quality of the 3-D percept. This lack of systematic and objective assessment of human performance might explain, at least in part, why some of the findings have been contradictory (see Dosher, Landy & Sperling, 1989 for review). The third problem associated with previous studies concerns the investigation of the number of views and features required to perceive SFM. A number of studies have shown that as few as 2–3 different views are sufficient to evoke 3-D percepts (Johansson, 1975; Borjesson & Von Hofsten, 1973; Lappin and Fuqua, 1983; Braunstein et al., 1987; Grzywacz, Hildreth, Inada & Adelson, 1988; Todd, Akerstrom, Reichel & Hayes, 1988), but there are contradictory reports on the effect of numbers of points on the *saliency* of SFM perception. Two groups report that SFM perception is improved by increasing the number of dots (Braunstein, 1962; Sperling, Landy, Dosher & Perkins, 1989), while one group of investigators claims that increasing the number of moving elements has no effect in their task (Todd et al., 1988). Another study contends that performance actually deteriorates (Braunstein et al., 1987). Similarly, although a larger number of frames appears to be helpful in creating more stable 3-D percepts (Wallach & O'Connell, 1953; White & Mueser, 1960; Green, 1961; Doner, Lappin & Perfetto, 1984; Braunstein et al., 1987; Grzywacz et al., 1988), it has also been reported that observers are able to perceive structure from as few as two different frames (Lappin, Doner & Kottas, 1980; Doner et al., 1984; Todd et al., 1988).

Another direction of research has attempted to find similarities between the performance of computational algorithms and recorded psychophysical performance. Grzywacz et al. (1988) demonstrated that perception of SFM builds up with increasing extent of angular rotation of a smoothly moving dot display. They interpret this finding as evidence for the use of the incremental rigidity scheme. This elegant position-based algorithm proposed by Ullman (1984) has been shown to perform best when the angular distance travelled by the points between the samplings is large and the stimulus is viewed for an extended period of time. However, three recent studies (Petersik, 1987; Todd et al., 1988; Mather, 1989) show that increasing the angular displacement of points between frames degrades perception of SFM.

In the present study, we use novel dynamic random-dot stimuli developed in our laboratory to investigate the spatiotemporal characteristics of human perception of SFM. We have two aims. The first is to delimit the minimal information required in both the spatial and temporal domains to evoke 3-D percepts from 2-D motion cues and to investigate how performance changes with manipulations of these stimulus parameters. In particular, we are interested in examining how the spatial and temporal factors interact. These interactions may help to explain inconsistencies between previous reports since many of these have employed only a few combinations of parameters. Several features of our stimulus minimize the problems associated with some previous investigations. The use of finite lifetimes leads to the reduction of shape cues in the display. It also proves a very powerful tool for investigating the temporal aspects of SFM perception and, together with the elimination of density cues from our displays, gives us rigorous control over stimulus parameters. In contrast to many other studies, we use a high display rate (70 Hz) and movies several seconds in length without cycling through the same set of frames. This allows for the more natural impression of continuous motion and prevents the subjects from memorizing the stimulus. Finally, our stimulus allows us to use reaction time and forced-choice experimental paradigms to assess perception of SFM quantitatively.

The second aim of our study is to investigate the kind of object representation generated in the SFM process. Investigations in depth perception using disparity information have demonstrated surface interpolation between feature elements (Collett, 1985; Morgan & Watt, 1982; Mitchison & McKee, 1985; Würger & Landy, 1989). Position-based SFM algorithms like Ullman's (1984) incremental rigidity scheme on the other hand generate wire-frame models from the visible features. These algorithms require the continuous presence of all features during the computation. This issue of whether object features have to be continuously present is of some biological significance because under natural viewing conditions objects

are often opaque and features rotate out of sight. In this study, we investigate whether surface representations, which do not require the continuous presence of all the points in the display, may be used for the perception of SFM.

GENERAL METHODS

Stimuli and protocol

The stimuli were dynamic random-dot displays presented on a CRT screen. The dots on the screen were the orthographic projection of points on the surface of a transparent, rotating cylinder. They lived for a pre-determined number of frames (finite lifetime) and appeared and disappeared asynchronously. These "flickering" displays minimize position cues in the stimulus since any configuration of dots will dissolve within a short time. In this way we attempt to examine the responses of the visual system to motion information alone.

If the average dot density is constant over the surface of a cylinder, its projection onto a two-dimensional surface yields an image with a greater density of dots at the edges representing the sides of the cylinder. In order to eliminate this density due, the random positions of dots are first generated in screen coordinates and then projected orthographically onto the surface of the cylinder. Since we repeat this procedure for every point when it gets replotted at the end of its lifetime there is always approximately equal density at all locations on the cylinder at any time.

The cylinder is rotated about a fixed, vertical axis. The parallel projection of the moving points is the display viewed by the observer on a CRT screen (Fig. 1). The horizontal velocity profile of the stimulus forms a half-cycle of a sinusoid between the two sides of the display while the velocity does not change along any vertical line. The stimulus points therefore

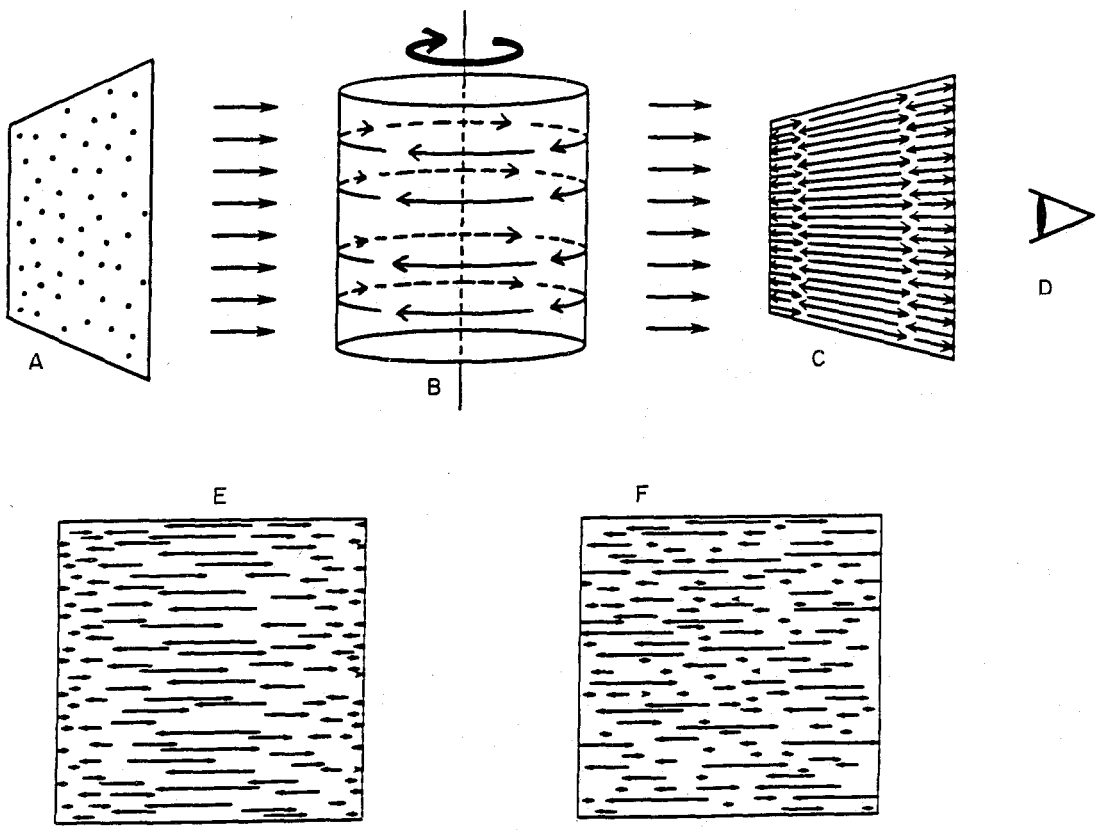


Fig. 1. (A–D) Visualization of the algorithm used to generate the stimulus. Points are randomly plotted onto a square. They are then projected orthographically onto a transparent cylinder which is rotated. The orthographic projection of each point is then stored in the memory of a PDP 11-73 computer. Subjects viewed the movies on a Hewlett–Packard 1311B CRT screen (phosphor P31). Viewing distance was 57 cm. The stimulus extended 6 × 6 angular deg. (E) Velocity field representing the structured stimulus. The velocity is small at the sides and high in the center, following a half cycle of a sinusoid along any horizontal line across the stimulus. (F) Velocity field representing the unstructured stimulus. To generate this stimulus every vector from the structured stimulus was offset by a random amount within the stimulus boundaries.

speed up while moving towards the middle of the display where the highest speeds are encountered and they slow down when heading towards the edges of the display.

Frames for the stimuli are computed off-line and entire movies are stored in the memory of the PDP 11-73 computer used for these experiments. The maximum possible length for each movie was 390 frames. The display rate of 70 Hz yielded movie lengths of 5.5 sec. Since the cylinder is transparent, half the points move in one direction and the other half in the opposite direction. The assignment of one direction of motion to the perceived front or rear surface of the cylinder is ambiguous and may reverse perceptually during viewing.

We refer to this cylinder stimulus as the "structured" display. Our "unstructured" stimulus is computed by randomly reshuffling all the vectors (i.e. the paths the dots travel in their lifetime) in the structured display (points falling off the edges of the display are wrapped around on the other side). Thus, the unstructured stimulus is made using the same set of vectors as in the structured case.* However, the structure of the velocity field (the distribution of velocity vectors across space) is altered. Perceptually, the unstructured stimulus appears to some observers as visual noise and to others as a cylinder filled with dots (the structured display being seen as a hollow one). Williams and Phillips (1986) report a similar observation when they constrain the directions present in a random dot pattern in which dots move in different directions.

In some experiments subjects were shown movies which contained a transition from the unstructured to the structured stimulus and were asked to detect the appearance of the rotating cylinder in a reaction-time task. In order to avoid artifactual cues, the transition between unstructured and structured displays is not an abrupt one. Operationally a frame number for transition is selected, however, a point always completes its pre-determined trajectory, even if it lives through the frame designated for transition. When such a point "dies" its new path is appropriate for the velocity field describing the structured stimulus. Thus, the transition is a period beginning with the designated frame for transition and completed within the point lifetime designated for the stimulus.

In other experiments subjects were required to discriminate between the unstructured stimulus and the rotating cylinder in a two-alternative forced-choice task. The two-alternative forced-choice task was employed whenever the stimulus duration was a crucial parameter.

In the reaction time paradigm each block of trials contained one movie which was identical to one of the others except it did not change to the structured stimulus (the "catch" trial). In this way we could gauge the approximate number of trials in which the observers obtained hits without a corresponding change in the structure of the display (i.e. the chance hit rate).

Depending upon the difficulty of the task, subjects were given a reaction time (RT) window starting after 300–500 msec and ending after 1500–2000 msec (0 msec referring to the onset of the transition period). If subjects responded within this window they received a short feedback tone—even if the trial was a catch one, since such trials were also allocated a randomly positioned RT window.† The change from the unstructured to the structured stimulus occurred randomly between 1 and 3 sec after the beginning of the movie. This proved sufficient to keep a low chance hit rate ($\sim$10–15%), i.e. the percentage of catch trials in which subjects responded within the RT window. In one set of experiments in which a very low number of points were used, the period before a change in structure was randomized over 1–7.7 sec. The reason for lengthening the fore-period for this more difficult task was that subjects guessed more, as indicated by the increased number of hits in the catch trials. To maintain the chance hit rate at 10–15%, the fore-period was lengthened. The frame rate was lowered to 35 Hz in these experiments.

The presentation of a movie ended with either the release of the key by the observer or the end of the RT window. A hit reflects the detection of the change from the unstructured to the structured display within the RT window

*It should be noted that this is different from the "no correlation" stimulus employed by Newsome and Paré (1988) and Downing and Movshon (1989). Rather than shuffling the motion *vectors* in the display these researchers introduce noise by randomly repositioning a certain percentage of the *points* between frames. In contrast to our procedure the average motion in their noise stimulus is therefore very different from the averaged motion in the signal stimulus.

†Since the "hit rate" on catch trials was generally lower than about 10–15% and only one movie in a block of about 10 was a catch, only in about 1% of all trials did the subjects receive such a "misleading" feedback.

(% hits = no. of correct responses/no. of correct + no. of late responses). The trials in which the subjects released the key before the RT window were not included in the computation of the hit rate. These early responses were not used in computing the hit rate to allow the use of different fore-periods, since they were more likely to occur in the tasks using longer fore-period ranges.

The experimental subjects viewed the display binocularly in a dimly lit room from a distance of 57 cm. Head movements were not restricted. The display subtended a visual angle of 6×6 deg and had a mean luminance of 1 cd m^{-2}. The size of the single points was about 0.5 mm (subtending 3 min visual arc) in diameter. A typical run would contain several movies which were presented in a random block design. The computer would randomly select the movies going through the whole set before a new cycle would begin. Typically, 100 stimuli were presented within a run, lasting about 10–15 min. After each run the movies were discarded and new ones were generated using new random number seeds. Thus the dot patterns were different for each block of trials. This avoided the possibility of subjects memorizing a movie.

The parameters of the stimulus systematically varied in this study were the number of points, the point lifetime, the angular rotation rate of the cylinder, the movie length, and how much of the stimulus was visible to the observer. For the conditions described above, the average 2-D velocity of a stimulus with an angular rotation rate of 35 deg/sec is 1.2 deg/sec. For a lifetime of 100 msec this corresponded to a path-length of 7 min visual angle. Two highly trained subjects (authors MH and ST) with corrected vision and no history of eye movement abnormalities were used. Since the subjects showed very similar performance all results presented here (except Fig. 2) are averages across them to smooth the curves and unclutter the figures. Whenever specific values are referred to in the text which are different for the two subjects their data are presented separated by a slash (/).

RESULTS

The use of limited point lifetimes shows a minimal temporal requirement

The most important feature of our stimulus, especially in comparison to previous work on the perception of SFM, is the finite point lifetime. Our first experiment investigated the influence of this parameter on the performance in the reaction time task. Each block of trials consisted of 10 movies (including the catch stimulus). Each of these movies had a different point lifetime ranging from 42 to 266 msec. The number of points was kept constant at 128; the angular rotation rate was 35 deg/sec.

The results for the two subjects are shown in Fig. 2. The psychometric curves demonstrate a minimal temporal requirement for the perception of SFM. Subjects cannot perform the task below a point lifetime of ~60 msec and the threshold (i.e. the point at which the curve reached 50% of its final height) is at ~69/81 msec. Peak performance is reached at a point lifetime of ~125 msec.

The threshold for the perception of SFM does not simply reflect a threshold of motion perception *per se* since subjects can clearly perceive the horizontal direction of motion in the stimulus when lifetimes where so short that no reliable percept of the 3-D object could be achieved. In other words, a longer integration time is needed to perceive SFM than just the direction of motion.

Threshold behavior with changes in point number and angular rotation rate

We investigated the influence of stimulus parameters on the threshold by varying the

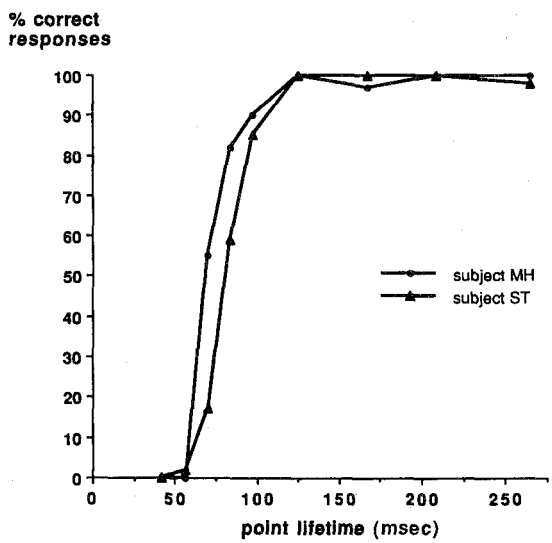


Fig. 2. Percent correct performance in reaction time task plotted as function of point lifetime. Each data point represents between 38 and 42 trials. The curves follow a sigmoidal shape. The threshold (50% of the final height) are at ~69/81 msec. Because of the similarity between the two subjects all the following graphs show their averaged responses.

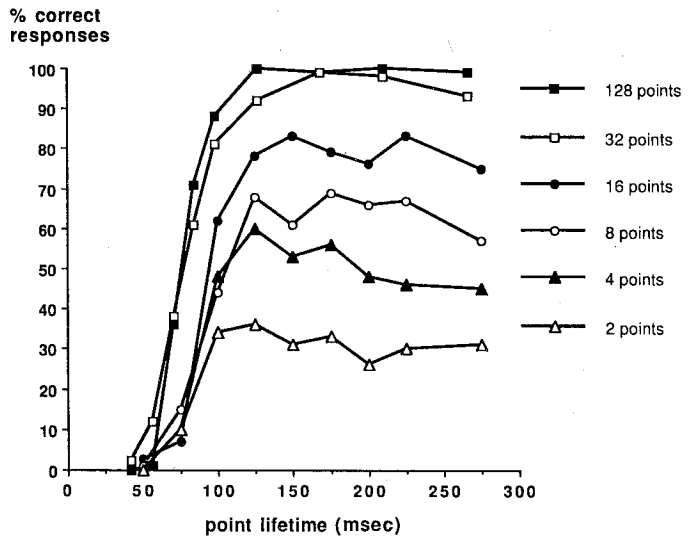


Fig. 3. Percent correct performance in reaction time task at different numbers of points. Note the similarity in thresholds compared to the wide variation in the heights of the plateaus.

number of points used for our stimuli between 2 and 128 points. The task and all other parameters were similar to the first experiment. Figure 3 shows the results. There is a marked decrease in performance when the number of points was lower than 32 but even at 4 points both subjects were still able to reach a performance of over 50%. Below saturation, a doubling of the number of points improved peak performance by ~16%. The threshold of all subjects' curves was between 75 and 90 msec and changed very little with large variations in the number of points and peak performance.

The next experiment studied the influence of the rotation rate (and therefore velocity in 2-D and 3-D) on performance. Figure 4a shows the results of our reaction time task. In contrast to the previous experiment all four curves have a very similar shape. The threshold was found to decrease somewhat with increasing rotation rate but only by about 40% (from ~49/51 msec at 140 deg/sec to ~81/87 msec at 21 deg/sec) over the seven-fold increase in rotation rate.

To investigate the possibility that improved performance at higher velocities reflects the increase in angular path-length we plotted performance for the different rotation rates against the angular path-length of the points. Figure 4b shows that the threshold for path-length was far from constant but rather increased by ~300% (from 1.7 to 6.5/7.0 deg) with the increase in speed.

An explanation for the threshold we observed comes from studies investigating optimal temporal properties of motion. We replotted the data from several such investigations reviewed by Nakayama (1985, his Fig. 6) in Fig. 4c together with our point lifetime thresholds (●) and the points where performance peaks in our task (■). Note the good correspondence over the wide range of tasks used in the different studies. We will argue in the Discussion below that these and other results suggest that precise measurements of velocity are important in perception of SFM.

Investigating build-up of performance

If indeed velocity measurements are employed for the perception of SFM the question arises as to how these measurements are used to compute 3-D shape. As in the previous experiments our use of finite point lifetimes proves to be an important tool to study how the visual system uses velocity measurements to compute SFM.

The current position-based algorithms, most notably Ullman's incremental rigidity scheme, sample the positions of a set of points in several discrete images and measure how the points change their positions relative to each other between these discrete images in order to compute their 3-D positions correctly. For such a scheme to work, all points have to be present during the entire viewing period so that their relative positions can be assessed. Two groups of investigators have already shown that the perception of SFM is possible with limited lifetimes (Todd et al., 1988; Dosher et al., 1989). But to truly test the biological validity of Ullman's algorithm, one has in addition to

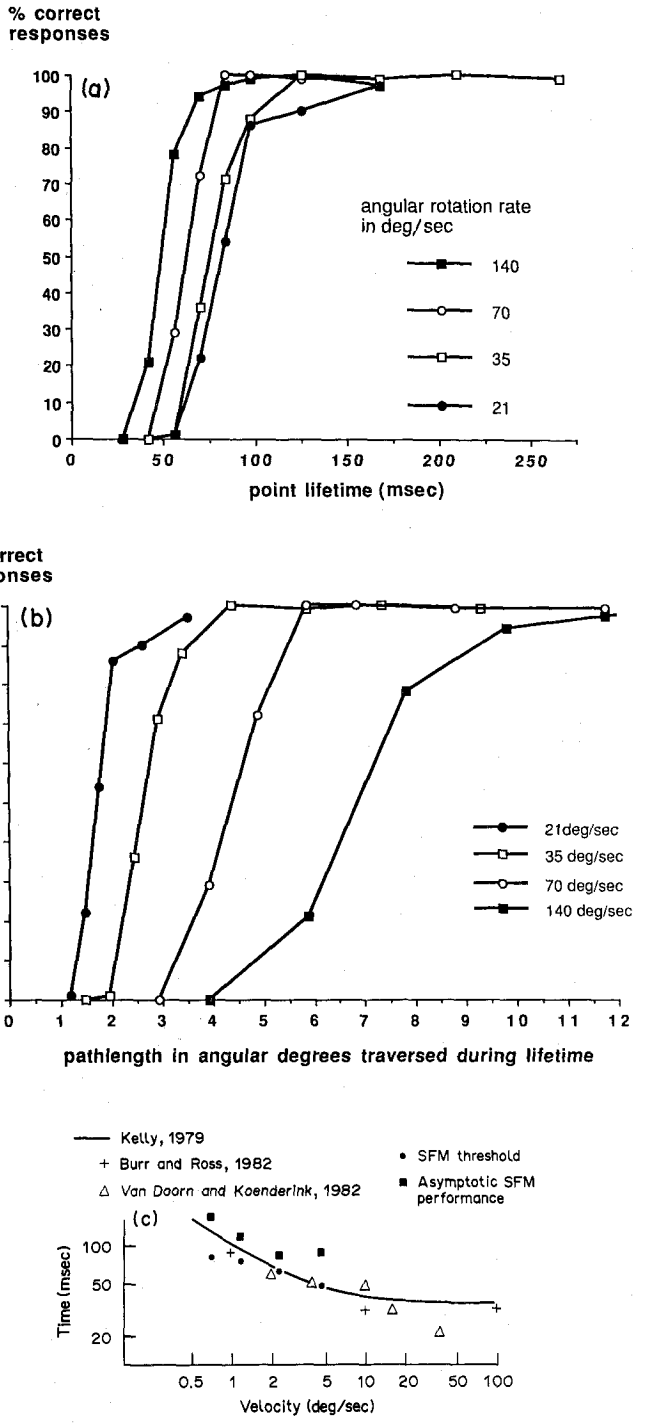


Fig. 4. (a) Percent correct response in reaction time task plotted as functon of rotation rates. Note the similarity in curve shapes. (b) Data from (a) replotted as function of pathlength in angular degrees travelled by individual points. Note the large shift in thresholds as compared to (a). (c) Threshold for SFM perception (●) and the lifetimes at which SFM performance peaks (■) as a function of 2-D velocity in the display plotted together with optimal temporal intervals for the inputs to hypothesized direction selective subunits as estimated by Nakayama (1985; his Fig. 6). The curve comes from calculating the optimal temporal intervals from the peak spatial and temporal frequency contrast sensitivity for detection (from the sine wave data of Kelly, 1979). (+)—from data measuring the contrast sensitivity for direction discrimination (Burr & Ross, 1982) analyzed in the same way as Kelly's data. (△)—derived from experiments requiring the observer to see coherent motion of random dots in a field dynamic visual noise (Van Doorn & Koenderink, 1982). It should be noted that despite wide differences in the experimental paradigm and observer task, estimates of optimal timing for a given velocity show considerable similarity.

show that stimuli exist for which the full SFM percept is *not* already achieved with just the information from one lifetime.

Figure 5 shows the reaction times for detecting the change from the unstructured stimulus to the rotating cylinder. They are plotted together with the respective point lifetimes of the stimulus. Two findings become immediately obvious: (1) reaction time varies as a function of point lifetime, ranging from ~1000 to ~700 msec for the point lifetimes tested; (2) the reaction times were always several times longer than the point lifetimes.

This second point attracted our attention because it suggests that the visual system is able to integrate information carried by points which appeared at different times. Unfortunately, it is not possible to determine definitively from these data how much of the reaction time is comprised of visual input and how much of it is computation time in the brain or motor reaction time. But such a measurement is crucial in the light of studies showing that SFM can, under certain circumstances, be perceived with just two frames (Lappin et al., 1980; Landy, Dosher, Sperling & Perkins, 1988). These studies seem to suggest that the brain might only need to view the stimulus for one point lifetime and then needs all the rest of the reaction time to process the information and execute the motor behavior. In order to investigate this issue subjects were shown stimuli of varying duration in a two-alternative forced-choice experiment. They were asked to report whether they saw a cylinder or unstructured stimulus. We varied the stimulus duration between 42 and 1680 msec and presented equal numbers of structured and unstructured stimuli. The lifetime was kept at 100 msec throughout this experiment. The results are plotted in Fig. 6a (○). Although subjects performed slightly above chance with a stimulus duration of one lifetime, there is a clear build-up in performance with increasing stimulus duration indicating that information is integrated over several point lifetimes.

One possible argument consistent with position-based approaches is that the visual system attempts to find a set of points whose lifetimes are aligned (in time) so that it can follow their composite pattern for several frames. Since our lifetimes are desynchronized it would be difficult to find such a set, especially since the number of points is large. To control for this possibility we ran an additional experiment which was identical to the previous one except that all the point lifetimes were synchronized so that they all began and ended their "lives" together. These results are also plotted in Fig. 6a (▲). The synchronization of lifetimes had no effect on the build-up of performance.

A way in which the visual system could perform the observed integration of information would be by fitting a surface through the data points (see Discussion for details). Such a surface interpolation scheme may only be used when the density of the points is high

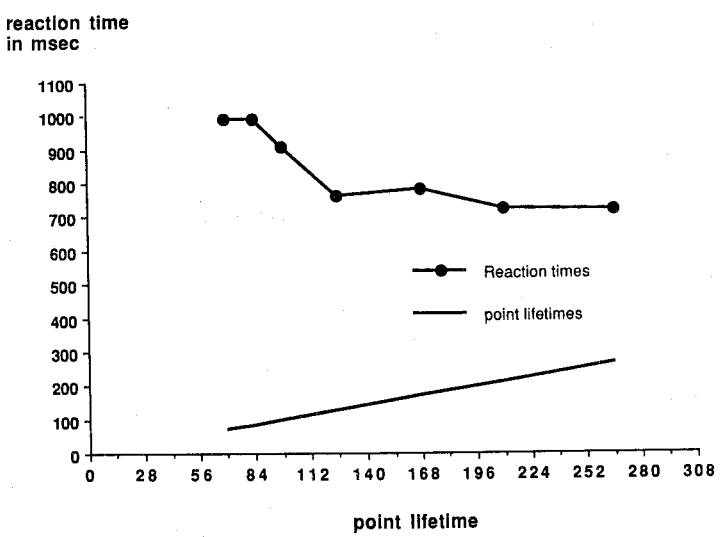


Fig. 5. Reaction time as a function of point lifetime. Point lifetime is also plotted to allow easy comparison with reaction time. Note that the reaction time is always many times longer than the point lifetime.

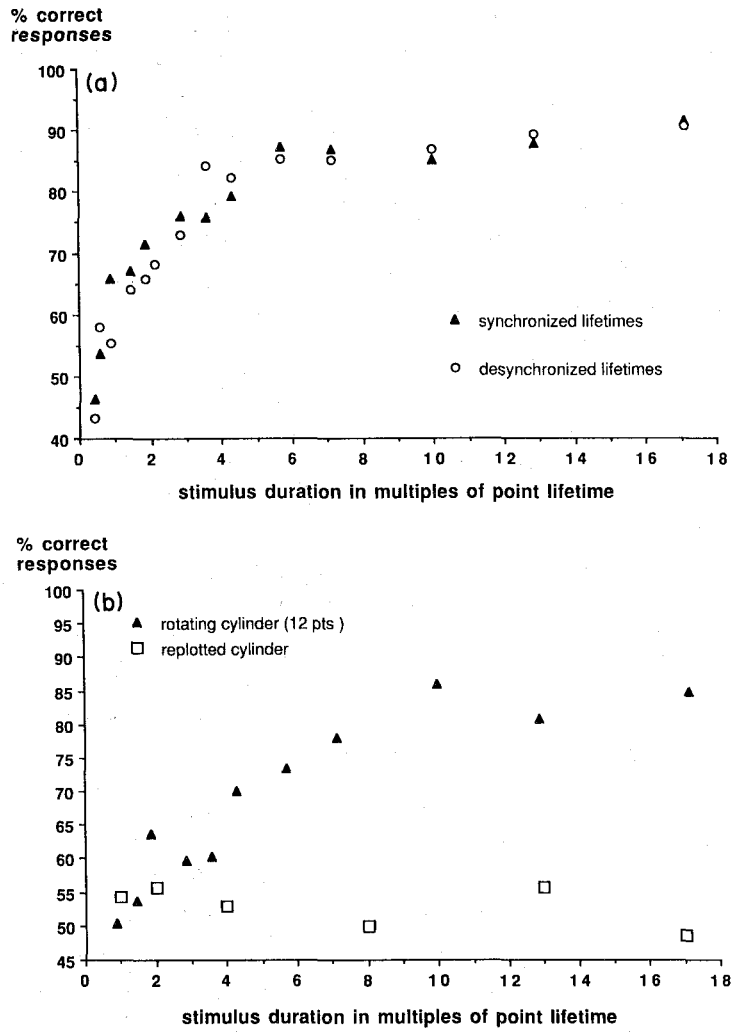


Fig. 6. (a) Percent correct responses in a two-alternative forced-choice task plotted as a function of stimulus duration. Note the long build up of performance. (b) (▲) percent correct responses using 12 points. (□) Percent correct when the points were repeatedly traveling along the same path.

and already closely approximates a surface. We therefore repeated our previous experiment with a cylinder composed of just 12 points. Fig. 6b (▲) shows the same phenomenon as the previous experiment with the buildup taking even longer. This result suggests that surface interpolation is also used in low density displays.

We performed another two-alternative forced-choice experiment to investigate if the observed build-up is due to factors other than the use of information from points which were widely separated in time. Movies were created with the same parameters as in the immediately preceding experiment except every point, after living through its first lifetime, was not randomly replotted but rather repositioned at its original location. It then moved through the same path as before, just to be repositioned at the original location, starting the cycle again. These movies contained the same number of points with the same point lifetime as used for the previous experiment but after the passage of the first point lifetime the stimulus contained no new information. Figure 6b (□) plots the results. It is obvious that the subjects are not able to perform the task, no matter how many lifetimes the stimulus was presented. In summary, we take the results from these experiments as strong evidence for the use of surface interpolation in the perception of SFM.

Global process in the perception of SFM

If SFM is perceived by making use of interpolated surfaces it is important to show that the tasks are performed using global rather than local cues. Due to the nature of our stimulus it is conceivable that subjects could

solve the tasks by monitoring local velocity coherence even though they were asked to use the perception of shape from motion as the only cue for responding in the task. In order to exclude the use of only local changes in velocity to perceive SFM, we ran a set of control experiments by masking out portions of the display.

In a reaction time task the subject (ST) was presented with movies in which most of the stimulus had been masked (this was achieved by simply not plotting the points falling within the mask) except for a square of 2 × 2 deg of visual angle in the centre. The visible area (~11% of the cylinder area) contained an average of 14 points. Since the percept of a full cylinder was obviously impossible, the subject was asked to respond to the appearance of two curved surfaces consistent with an interpretation of a partial view of a rotating cylinder. The results are plotted in Fig. 7a in comparison to the performance achieved using a full size stimulus with either the same number of points or the same dot density. As is apparent from the data in Fig. 7a it is nearly impossible to perform the task when the mask is present.

We performed another set of experiments to investigate the influence of the size, shape and position of the mask. In these experiments only 25% of the stimulus was masked away. Sketches of the different masks are shown in Fig. 7b. The two types of vertical masks cover the edges or the center of the display. The horizontal masks also cover the center or the edges but the areas they occlude are redundant because all velocities are still represented in the unmasked areas of the cylinder. The horizontal masks were included to control for the effect of disrupting the stimulus by breaking it into two parts (central mask) and for the effect of decreasing overall stimulus size (central and peripheral mask). The results are plotted in Fig. 7c. The data for the peripheral vertical mask (●) show a marked reduction in performance in comparison to the horizontal peripheral mask (■). The results using the central vertical mask show an even stronger reduction in performance (○). Performance with the horizontal masks was similar to that measured with an unmasked display (□ and ■ compared to dotted line).

GENERAL DISCUSSION

In these experiments we have attempted to examine the spatiotemporal characteristics of the perception of SFM. We used a reaction time task in which subjects were asked to detect a change in the structure of the presented stimulus and several two-alternative forced-choice paradigms.

We showed:

(1) there is a point lifetime threshold for detection of SFM which remains fairly constant (50–85 msec) over a wide range of number of points and velocities, although it does increase somewhat with decreasing angular rotation rates (Figs 2–4);

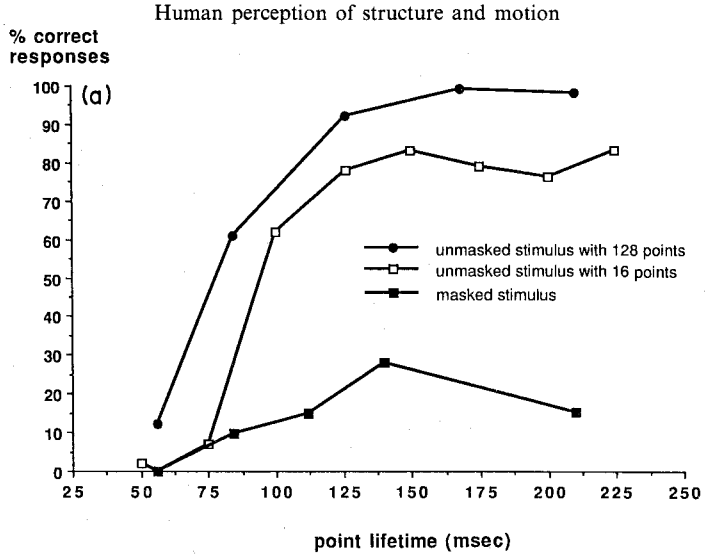
(2) this threshold reflects a minimal temporal requirement and not a minimal path-length (or threshold for detection of motion) (Fig. 4);

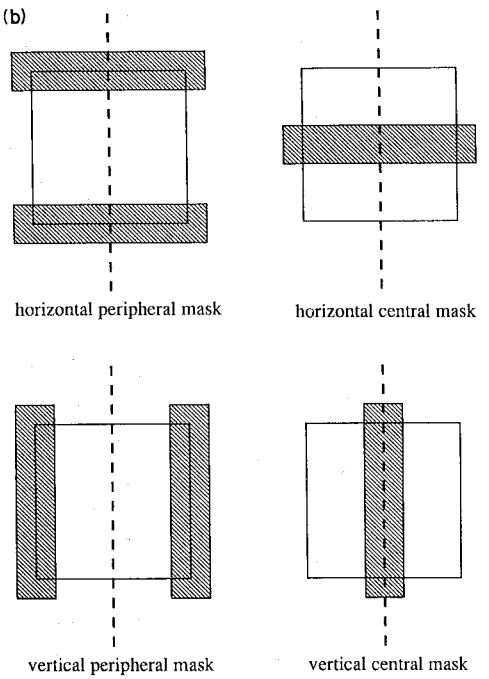
(3) reaction times (RTs) for perceiving SFM are long (Fig. 5), presumably reflecting a process which integrates information temporally across several points lifetimes (Fig. 6). This process is global as shown by the effect of even a small mask on performance (Fig. 7).

Temporal characteristics

One of the most interesting findings of our study is the demonstration of a point lifetime threshold for the detection of SFM. The sharp drop in performance for point lifetimes shorter than 50–85 msec indicates a minimal temporal requirement. Why does the visual system need point lifetimes of at least this duration? And

Fig. 7 (*Opposite*). Performance in reaction time task comparing unmasked stimuli with masked stimuli. The number of points in the unmasked part of the stimulus averaged 32 to allow easy comparison with our data from the unmasked stimulus. (a) Performance when only ~11% of the stimulus was visible (average of ~14 points) compared to performance using full-size stimulus with either same number of points (□) or with the same dot density (●). (b) Shape and location of masks used to cover 25% of the stimulus. The two horizontal masks cover redundant areas of the stimulus while the vertical mask cover nonredundant areas. (c) Comparison of the effect of a 25% mask of different shape and orientation. (○) Performance using central vertical mask; (□) performance using central horizontal mask; (●) performance using peripheral vertical mask; (■) performance using peripheral horizontal mask; (- - - -) performance with the unmasked cylinder using 32 points (from Fig. 3). Note the difference in performance between stimuli in which a nonredundant part was masked away (circles) and when a redundant part of the stimulus was invisible (squares).


% correct responses
(a)
100
90
80
70
60
50
40
30
20
10
0
25
50
75
100
125
150
175
200
225
250
unmasked stimulus with 128 points
unmasked stimulus with 16 points
masked stimulus
point lifetime (msec)


(b)
horizontal peripheral mask
horizontal central mask
vertical peripheral mask
vertical central mask

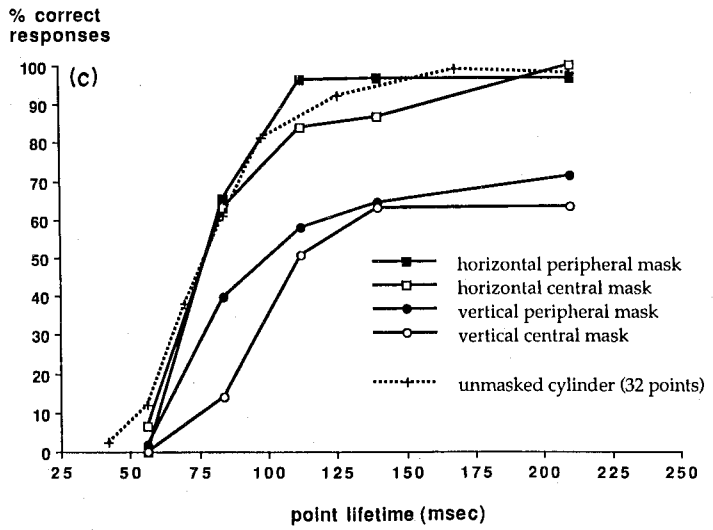

% correct responses
(c)
100
90
80
70
60
50
40
30
20
10
0
25
50
75
100
125
150
175
200
225
250
horizontal peripheral mask
horizontal central mask
vertical peripheral mask
vertical central mask
unmasked cylinder (32 points)
point lifetime (msec)

Fig. 7 (a–c)

why does this required time fall with increasing angular rotation rate?

One explanation might be that in order to compute the 3-D locations of points, the visual system samples their 2-D positions in several discrete images and measures how these change relative to each other between frames to compute their 3-D locations. The observed threshold might therefore be attributed to a requirement to sample a minimum number of images in order to assign 3-D positions correctly. Ullman (1984) has in fact proposed such a position-based algorithm—the incremental rigidity scheme—which over a number of images correctly assigns 3-D locations to every point. The observed threshold could therefore represent the minimum point lifetime that is needed for the visual system to obtain enough samples of the positions of the points. Note that for such a scheme, velocity measurements are not required. In fact, Ullman's scheme does not put any restrictions on the sampling frequency or even the order of frames, therefore making even implicit velocity calculations effectively impossible.

The performance of this position-based scheme depends on the accuracy with which the displacements of points between frames is assessed. It is therefore highly sensitive to errors in measuring 2-D positions if the displacements between sampled images are small. As these displacements becomes smaller, the algorithm requires larger and larger amounts of rotation to approximate the correct solution (for a detailed analysis of this problem see Grzywacz & Hildreth, 1987). Indeed, Grzywacz and Hildreth have shown that a continuous implementation of the incremental regidity scheme is highly unstable. These authors therefore suggest that continuous motion is discreetly sampled (to obtain position) and the actual motion between the samples is used only to solve the correspondence problem between points in the images.

The preceding two paragraphs appear to offer a possible explanation for the minimal temporal requirement as well as its shift with increasing rotation speed: position-based approaches such as Ullman's need a set of discrete images of a point but the exact number of views might not be critical as long as a minimum overall path-length (and therefore the overall angular extent of rotation) is inspected. However, Fig. 4b shows that no speed-independent path-length threshold exists. Position based algorithms would have predicted that the extent threshold would be relatively constant since extent of the movement (and not velocity) is the important parameter for these algorithms. Grzywacz and Hildreth's (1987) analysis points to another important problem: Fig. 4b shows that subjects in our experiments reach peak performance with only a few degrees of angular rotation for each point. Even disregarding the finite lifetimes, the rotation of the whole cylinder was only about 30 deg before the subjects responded. The analysis presented by Grzywacz and Hildreth (1987, see their Fig. 2, and personal communication) shows that the incremental rigidity scheme is unable to perform accurately with such small amounts of rotation. Thus, it seems unlikely that the sampling requirements of a position-based algorithm such as Ullman's can account for the threshold or the high level of performance observed in our experiments. The need to track object features over extended rotation angles points to another problem of algorithms such as Ullman's: they are vulnerable to occlusions and rotation out of sight when objects are opaque.

It is useful to point out that occlusions can occur in two different domains. In the object domain the occluder hides part of the object (thereby also limiting lifetimes). Our masking is an example. The other type of occlusion occurs at the level of the features; an example would be when a feature on an opaque rotating object rotates out of sight. Our use of limited lifetimes is an example of feature based occlusion. Our experimental results clearly demonstrate that the visual system is able to cope with feature based occlusions whereas algorithms which have to track features continuously fail.

An alternative means of solving the SFM problem is to use velocity information rather than the raw positions of points in each image. McKee and Welch (1985) showed that subjects need to view a moving bar for 80–100 msec for asymptotic velocity discrimination. Because this range of point lifetime corresponds very closely to the threshold observed using our SFM stimuli we suggest that the point lifetime thresholds observed in our experiments reflects the time required to measure velocity accurately. Our explanation for the threshold, therefore, is that subjects use velocity measurements to solve the SFM problem and simply need to see points for a minimum time before they can correctly measure their 2-D velocities. This

interpretation of our data is strengthened by findings from several laboratories which show that increased velocities allow the same level of performance with shorter stimulus durations (McKee & Welch, 1985; De Bruyn & Orban, 1988) and that optimal temporal displacements in apparent motion sequences are shorter for faster velocities (Nakayama, 1985). Figure 4a shows that perception of SFM has very similar characteristics and the correspondence with data from visual motion experiments is further emphasized in Fig. 4c. This shows how the data presented here compares with the spatiotemporal characteristics of velocity perception (from Nakayama, 1985). The requirement for accurate velocity measurement offers an explanation for our observation that stimuli could be perceived as moving at lifetimes too short for reliable SFM perception. The detection of motion alone is therefore simply not sufficient for perception of SFM; and the somewhat longer lifetime needed to measure velocities accurately is what leads to the difference.

Support for this velocity hypothesis comes from the finding that lesions of area MT, a region in primate visual cortex which contains neurons tuned to stimulus velocity (Maunsell & Van Essen, 1983), have been found to impair perception of SFM (Siegel & Andersen, 1986). The results of two recent psychophysical studies are also in agreement with this hypothesis. Mather (1989) has shown that SFM depends on the outputs of low-level or "short-range" motion detectors and Dosher et al. (1989) argue against position-based algorithms because their subjects showed only a weak loss of performance if the point lifetime was reduced from 30 to 2 or 3 frames. This finding might seem surprising in light of the strong effects of point lifetime in this range in our study but two factors might explain this discrepancy. First, Dosher et al. use very long stimulus onset asynchronies (SOAs) and as a result their shortest point lifetimes were 133 and 200 msec. Their slight decline in performance for 2 frames might therefore represent the high end of our SFM threshold. Second, it is not so surprising that performance improves so little from 2 to 30 frame lifetimes in the light of a recent study by Snowden and Braddick (1989). These investigators showed that long SOAs severely inhibit the ability to use temporal recruitment (Nakayama & Silverman, 1984) to improve performance in long frame sequences as compared to short frame sequences in a direction of motion task. This effect is so pronounced for SOAs between 50 and 100 msec that performance peaks after only about four displacements. These data suggest that Dosher et al. observed such a small improvement with longer lifetimes (i.e. larger frame sequences) because their choice of SOAs effectively prevents the visual system from making use of the more numerous frame sequences. Also they used conditions in which performance was high even for short lifetimes and this may also account for why they saw so little improvement.

Surface interpolation and spatiotemporal integration

How might velocity measurements be used to compute 3-D shape? Several algorithms which require the measurement of velocity (or higher derivatives) have been proposed (Clocksin, 1980; Longuet-Higgins & Prazdny, 1980; Hoffman, 1982; Koenderink and Van Doorn, 1986). Recently, we suggested that a means of solving the SFM problem is to measure the velocities of as many points as possible across the surface of the object, fit a smooth 2-D velocity field to the measurements, and use this velocity map to derive a 3-D surface representation of the object (Husain, Treue & Andersen, 1989). In theory, such a representation may also be computed by fitting a smooth surface through the 3-D positions derived from the 2-D velocities of each point. These 3-D positions could only be assigned after comparing velocities across the stimulus since depth is determined by relative velocities between different parts of the stimulus rather than absolute local velocities. If such a global comparative process is already performed at the 2-D level it seems more parsimonious that the smoothing and the interpolation of the velocity field is also done at the same time. On the other hand the process of surface interpolation might operate on the 3-D positions to allow for cue-integration especially for stereopsis. Surface interpolation offers a more plausible solution to the SFM problem than current position-based algorithms given that the visual system evolved in an environment where tracked individual features frequently are only present for a short period of time.

In the experiments reported here, we found that RTs for detecting the change from unstructured to structured stimuli were remarkably long, ranging from 700 to more than 1000 msec.

These values are much longer than the point lifetimes needed for perceiving SFM and suggest that the visual system samples the stimulus for several times longer than the lifetime of any one set of points. This observation is consistent with the scheme outlined above: the system integrates measurements from several sets of points to compute a reasonably accurate surface representation.

Our two-alternative forced-choice experiment demonstrates that the long RTs reflect visual input of several lifetimes and not just a long computation or motor response time (Fig. 6b). Such behavior does not exclude the possibility that a position-based algorithm may be employed, but clearly it cannot be explained by algorithms such as Ullman's incremental rigidity scheme. Furthermore, the results support the surface interpolation hypothesis described above. The possibility that the observed build-up in performance is not due to actual integration of information over time but might just reflect an unrelated intrinsic phenomenon of the visual system is countered by our control experiment in which points were replotted so that they moved through the same paths over and over again. Performance did not increase above chance (Fig. 6b).

There is no *a priori* reason why the visual system should not improve its performance when repeatedly presented with the same set of frames. It may be that subjects are not able to use fully the information presented in the first few frames, especially if the stimulus appears on an otherwise dark and featureless screen. This is especially true if the number of different frames is small and the motion measurement across space is not totally in parallel but involves some "patch by patch" measurements. It has previously been observed "that the perception of rigid rotation from two-frame sequences may be critically dependent on a repetitive oscillation of the display" when high-density stimuli were used (Todd et al., 1988). In our displays, when 12 points were made to retrace their steps over and over again, they probably carry so little information that it can be measured within the first lifetime. In agreement with these considerations we observe some build-up in performance when using high number of points. Under better conditions (i.e. if long point lifetimes, high number of points, and higher rotation rates are used) good performance can already be achieved within the first point lifetime (as also seen by Landy et al., 1988).

It could be argued that our task, since it does not require the subjects to distinguish between two different structures, can be solved by just measuring local coherence of the velocity field. This seems unlikely given that our subjects were asked to use the overall shape of the stimulus as the cue for their response. But stronger experimental evidence comes from our masking experiment. The results show that occluding the cylinder by as little as 25% leads to a marked reduction in performance (Fig. 7a). This is only the case if the mask covers nonredundant areas of the cylinder. If the same size mask is placed horizontally performance is not reduced. This control rules out effects of the number of dots or their density. If subjects monitor coherence locally the observed effect would not be expected since they could easily shift their attention to any unmasked region of the cylinder. The observed reduction of performance with the peripheral mask cannot be due to the subject's preferred monitoring of the stimulus edges since the central vertical mask leads to an even stronger reduction in performance. This result would be expected with a surface interpolation algorithm since it would have problems interpolating across the central mask while at the edges where the velocities are already low the velocity field would be smoothly interpolated to the stationary surround.

Our use of limited lifetimes turns out to be critical in ruling out another possible explanation of the above results. Had we used unlimited lifetimes one could have argued that the performance with the vertical mask was degraded because individual point paths are cut short by the mask and therefore are less visible and poolable. But given that the vertical peripheral mask leaves more than 120 deg of each surface of the cylinder visible and the points travel only through about 2–7 deg (depending on the lifetime used) the above argument does not apply.

Spatial characteristics

The surface scheme depends critically upon integrating samples taken at different positions (in space and time) across the surface of the object. It predicts that performance should improve as the number of samples taken per unit time increases. In agreement with this prediction, we found that (for the range 2–32 points) peak performance improved with increasing number of points (Fig. 3).

Our results are not in agreement with those of Braunstein et al. (1987) who showed that increasing the number of points between 2 and 5 actually worsened perception of SFM. The most likely reason for this discrepancy lies in the differences between the two tasks. Braunstein et al. noted that: "The theoretical analyses considered in the present study were concerned with recovering depth coordinates for individual points. The task that we used would not be appropriate for studying analyses concerned with recovering surface structure. Different results might be expected for number of points if the task involved detection of surfaces or discrimination among surfaces." It is of interest to note that Braunstein and his colleagues report that their subjects were able to see structures when just one frame of movie was displayed, suggesting that pattern information as well as motion was available as cues in their displays.

Our data demonstrate clearly, as others have done previously, that it is possible to see structure with some degree of reliability with only a few points. This is to be expected from the surface scheme outlined above since, provided the viewing time is long enough, the spatial sampling will be sufficiently dense (due to temporal integration) to compute a surface representation.* This suggestion of a trade-off between viewing time and number of points is strengthened by our results presented in Fig. 6a and b. These data show that the visual system integrates over a longer period of time before it reaches peak performance if fewer points are presented. What is not expected *a priori* from the surface hypothesis, however, is the finding that peak performance fails to reach 100% correct responses when low numbers of points (2–16) are used since, in principle, there should be sufficient data present to extract the 3-D structure given a long enough stimulus duration. This result therefore suggests a limited spatio-temporal memory capacity for the SFM system.

Algorithms and motion transparency

The proposal that the visual system integrates many samples over space and time to compute a 3-D surface representation of the object has also been advanced to account for perception of short range coherent motion (Van Doorn & Koenderink, 1984; Snowden & Braddick, 1989). A large number of algorithms for 2-D velocity measurement have been proposed which perform some velocity integration, averaging or smoothing (Hildreth & Koch, 1987; Horn & Schunk, 1981; Zucker & Iverson, 1986; Yuille & Grzywacz, 1988; Bülthoff, Little & Poggio, 1989) over patches of measured velocities to compute a smooth map of velocity over space. Some of these algorithms have also been implemented in neural networks (Wang, Mathur & Koch, 1989). These are attractive schemes since they employ techniques which can account for a number of perceptual phenomena, e.g. motion capture (Ramachandran & Anstis, 1983) and the aperture problem (Wallach, 1976; Marr & Ullman, 1981).

Unfortunately, none of these algorithms can deal with the recovery of SFM for transparent objects such as our rotating cylinder: vectors (with opposing direction) from the front and rear surface are assigned to one surface, and the averaging of velocities over a patch yields zero velocity. Evidently, an additional requirement for the successful application of these algorithms is the segregation of surfaces prior to smoothing. Recent work from our laboratory has demonstrated that many cells tuned for direction of motion in monkey striate cortex act as simple directional filters which are not influenced by the presence of dots moving in the nonpreferred direction (Erickson, Snowden, Andersen & Treue, 1989). Thus as early as V1 two transparent surfaces moving in opposite directions will excite different populations of neurons, thereby providing the first step for surface segregation based on direction of motion. Presumably, similar mechanisms allow for surface segregation based on speed as shown psychophysically in several studies (Ramachandran, Cobb & Rogers-Ramachandran, 1988; Andersen, 1989).

In conclusion, we have demonstrated the existence of a rather invariant point lifetime threshold for perception of 3-D structure-from-motion. Our data suggest that this threshold reflects the limits of accurate velocity perception and that accurate velocity measurements are critical for the computation of structure from motion. Furthermore, our results support the hypothesis that the visual system integrates information over space and time by computing

*This is also our explanation for the ability to perceive at least a very crude perception of SFM with as little as two points. Interestingly the percept of such a sparse stimulus at the short lifetimes used in our experiment is one containing about 5–10 rather than just two points which is additional perceptual evidence for temporal integration.

a 3-D surface representation of the object. Such a process renders the visual system more flexible in its use of transient features for the SFM computation than current position-based schemes and is therefore more plausible. Finally, our data suggest that both velocity measurements and surface interpolation reflect global processes.

Acknowledgements—We would like to thank Ken Nakayama for allowing us to use his figure. We are grateful to Shabtai Barash, Noberto Grzywacz, Ellen Hildreth and Robert Snowden for helpful discussions. This work was supported by grants to RAA from the NIH, the Sloan Foundation, the Whitaker Health Sciences Foundation and the Educational Foundation of America. S.T. is a Fellow of the Evangelisches Studienwerk Villigst, F.R.G. and is supported by the Educational Foundation of America and M.H. is a Harkness Fellow.

REFERENCES

Andersen, G. J. (1989). Perception of three-dimensional structure from optical flow without locally smooth velocity. *Journal of Experimental Psychology: Human Perception and Performance, 15*, 363–371.

Borjesson, E. & Von Hofsten, C. (1973). Visual perception of motion in depth: Application of a vector model to three-dot motion patterns. *Perception and Psychophysics, 13*, 169–179.

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., 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.

Bülthoff, H., Little, J. & Poggio, T. (1989). A parallel algorithm for real-time computation of optical flow. *Nature, London, 337*, 549–553.

Burr, D. C. and Ross, J. (1982). Contrast sensitivity at high velocities. *Vision Research, 22*, 479–484.

Clocksin, W. F. (1980). Perception of surface slant and edge labels from optical flow: A computational approach. *Perception, 9*, 253–269.

Collett, T. S. (1985). Extrapolating and interpolating surfaces in depth. *Proceedings of the Royal Society, London B, 224*, 43–56.

De Bruyn, B. & Orban, G. A. (1988). Human velocity and direction discrimination measured with random dot patterns. *Vision Research, 28*, 1323–1335.

Doner, R., 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.

Dosher, B. A., Landy, M. S. & Sperling, G.(1989). Kinetic depth effect and optic flow. I. 3D shape from Fourier motion. *Vision Research, 29*, 1789–1813.

Downing, C. & Movshon, J. A. (1989). Spatial and temporal summation in the detection of motion in stochastic random dot displays. *Investigative Ophthalmology and Visual Science (Suppl.), 30*, 72.

Erickson, R. G., Snowden, R. J., Andersen, R. A. & Treue, S. (1989). Directional neurons in awake rhesus monkeys: Implications for motion transparency. *Society for Neuroscience Abstracts, 15*, 323.

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). Incremental rigidity scheme for recovering structure from motion: Position-based versus velocity-based formulations. *Journal of the Optical Society of America, 4*, 503–518.

Grzywacz, N. M., Hildreth, E. C., Inada, V. K. & Adelsen, E. H. (1988). The temporal integration of 3-D structure from motion: A computational and psychophysical study. In Von Seelen, W., Shaw, G. & Leinhos, U. M. (Eds), *Organization of neural networks* (pp. 239–259). Weinheim: VCH.

Hildreth, E. C. & Koch, C. (1987). The analysis of visual motion: From computational theory to neuronal mechanisms. *Annual Review of Neuroscience, 10*, 477–533.

Hoffman, D. D. (1982). Inferring local surface orientation from motion fields. *Journal of the Optical Society of America, 72*, 888–892.

Horn, B. K. P. & Schunk, B. G. (1981). Determining optical flow. *Artificial Intelligence, 17*, 185–203.

Husain, M., Treue, S. & Andersen, R. A. (1989). Surface interpolation in 3-D structure-from-motion perception. *Neural Computation, 1*, 324–333.

Johansson, G. (1975). Visual motion perception. *Scientific American, 232*, 76–88.

Kelly, D. H. (1979). Motion and vision. II. Stabilized spatio-temporal threshold surface. *Journal of the Optical Society of America, 69*, 1340–1349.

Koenderink, J. J. & Van Doorn, A. J. (1986). Depth and shape from differential perspective in the presence of bending deformations. *Journal of the Optical Society of America, A3*, 242–249.

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 and Cognition, 88-4*. NYU Report Series.

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

Lappin, J. S., Doner, R., & Kottas, B. L. (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, London B, 208*, 385–397.

Marr, D. & Ullman, S. (1981). Directional selectivity and its use in early visual processing. *Proceedings of the Royal Society, London B, 211*, 151–180.

Mather, G. (1989). Early motion processes and the kinetic depth effect. *Quarterly Journal of Experimental Psychology 41A*, 183–198.

Maunsell, J. H. R. & Van Essen, D. C. (1983). Functional properties of neurons in the middle temporal visual area (MT) of the macaque monkey. I. Selectivity for stimulus direction, speed and orientation. *Journal of Neurophysiology, 49*, 1127–1147.

McKee, S. P. & Welch, L. (1985). Sequential recruitement in the discrimination of velocity. *Journal of the Optical Society of America, A2*, 243–251.

Miles, W. R. (1931). Movement interpretations of the silhouette of a revolving fan. *American Journal of Psychology, 43*, 392–405.

Mitchison, G. J. and McKee, S. P. (1985). Interpolation in stereoscopic matching. *Nature, London, 315*, 402–404.

Morgan, M. J. & Watt, R. J. (1982). Mechanisms of interpolation in human spatial vision. *Nature, London, 299*, 553–555.

Nakayama, K. (1985). Biological image motion processing: A review. *Vision Research, 25*, 625–660.

Nakayama, K. & Silverman, G. H. (1984). Temporal and spatial characteristics of the upper displacement limit for motion in random dots. *Vision Research, 24*, 293–299.

Newsome, W. T. & Paré, E. B. (1988). A selective impairment of motion perception following lesions of the middle temporal visual area (MT). *Journal of Neuroscience, 8*, 2201–2211.

Petersik, T. J. (1987). Recovery of structure from motion: Implications for a performance theory based on the structure-from-motion theorem. *Perception and Psychophysics, 42*, 355–364.

Poggio, T. & Koch, C. (1985). Ill-posed problems in early vision: From computational theory to analog networks. *Proceedings of the Royal Society, London B, 226*, 303–323.

Ramachandran, V. S. & Anstis S. M. (1983). Displacement thresholds for coherent apparent motion in random-dot patterns. *Vision Research, 23*, 1719–1724.

Ramachandran, V. S., Cobb, S. & Rogers-Ramachandran, D. (1988). Perception of 3-D structure from motion: The role of velocity gradients and segmentation boundaries. *Perception and Psychophysics, 44*, 390–393.

Siegel, R. M. & Andersen, R. A. (1986). Motion perceptual deficits following ibotenic acid lesions of the middle temporal area in the behaving rhesus monkey. *Society for Neuroscience Abstracts, 12*, 1183.

Siegel, R. M. & Andersen, R. A. (1988). Perception of three-dimensional structure from motion in monkey and man. *Nature, London, 331*, 259–261.

Snowden, R. J. & Braddick, O. J. (1989). The combination of motion signals over time. *Vision Research, 29*, 1621–1630.

Sperling, 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 and Performance, 15*, 826–840.

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 and Psychophysics, 43*, 179–188.

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

Van Doorn, A. J. & Koenderink, J. J. (1982). Temporal properties of the visual detectibility of moving spatial white noise. *Experimental Brain Research, 45*, 179–182.

Van Doorn, A. J. & Koenderink, J. J. (1984). Spatiotemporal integration in the detection of coherent motion. *Vision Research, 24*, 47–53.

Wallach, H. (1976). On perceived identity: 1. The direction of motion of straight lines. In Wallach, H. (Ed.), *On perception* (pp. 201–216). New York: Quadrangle.

Wallach, H. & O'Connell, D. N. (1953). The kinetic depth effect. *Journal of Experimental Psychology, 45*, 205–217.

Wang, H. T., Mathur, B. & Koch, C. (1989). Computing optical flow in the primate visual system. *Neural Computation, 1*, 92–103.

White, B. W. & Mueser, G. E. (1960). Accuracy in reconstructing the arrangements of elements generating kinetic depth displays. *Journal of Experimental Psychology, 60*, 1–11.

Williams, D. & Phillips, G. (1986). Structure from motion in a stochastic display. *Journal of the Optical Society of America, A3*, P12.

Würger, S. M. & Landy, M. S. (1989). Depth interpolation with sparse disparity cues. *Perception, 18*, 39–54.

Yuille, A. L. & Grzywacz, N. M. (1988). A computational theory for the perception of coherent visual motion. *Nature, London, 333*, 71–74.

Zucker, S. W. & Iversen, L. (1986). From orientation selection to optical flow. *Computer Vision, Graphics and Image Processing, 37*, 196–222.