Pergamon

0042-6989(94)E0069-W

*Vision Res.* Vol. 35, No. 1, pp. 139–148, 1995

0042-6989/95 $7.00 + 0.00

# Structure-from-motion: Perceptual Evidence for Surface Interpolation

STEFAN TREUE,*† RICHARD A. ANDERSEN,*‡ HIROSHI ANDO,§ ELLEN C. HILDRETH‖¶

*Received 8 April 1993; in revised form 3 March 1994*

**Dynamic random-dot displays representing a rotating cylinder were used to investigate surface interpolation in the perception of structure-from-motion (SFM) in humans. Surface interpolation refers to a process in which a complete surface in depth is reconstructed from the object depth values extracted at the stimulus features. Surface interpolation will assign depth values even in parts of the object that contain no features. Such a "fill-in" process should make the detection of featureless stimulus areas ("holes") difficult. Indeed, we demonstrate that such holes in our rotating cylinder can be as wide as one-quarter of the stimulus before subjects can reliably detect their presence. Subjects were presented with a variation on the rotating cylinder in which all dots were oscillating either in synchrony or asynchronously. Subjects perceive a rigidly rotating cylinder even when such a percept is not in agreement with the physical stimulus. To reconcile this discrepancy between actual and perceived stimulus we propose that individual points contribute to a surface based object representation and that in this process the visual system looses access to the identity of the individual features that make up the surface. Finally we are able to explain a variety of previously documented perceptual peculiarities in the perception of structure-from-motion by arguing that the perceptual interpretation of the object's boundaries influences the surface interpolation process. These findings offer strong perceptual evidence for a process of surface interpolation and are also physiologically plausible given results from recordings in awake behaving monkey cortical areas V1 and MT. The companion paper demonstrates how such a surface interpolation process can be incorporated into a structure-from-motion algorithm and how object boundaries can influence the perception of structure-from-motion as has been demonstrated before and in this paper.**

Structure-from-motion *Surface interpolation* Spatial integration Temporal integration Random-dot patterns

## INTRODUCTION

When the visual system is trying to recover the three-dimensional (3-D) shape of an object it is presented with the problem of only having access to the two-dimensional (2-D) images projected onto its retinae. Besides the disparity between the images in the two eyes a multitude of monocular cues (like shading, texture, occlusions, and motion) enable the visual system to perform the task despite this limitation. Extracting the 3-D shape of objects from the relative 2-D motion of their parts is called structure-from-motion (SFM) or the kinetic depth effect and has been studied extensively since its first description by Miles (1931).

While the original studies of SFM were performed using shadows of real objects as stimuli (Miles, 1931; Wallach & O'Connell, 1953), the more recent stimuli of choice have been computer generated moving random-dot patterns. These displays represent the projection of rotating objects covered with dots. The advantage of such stimuli is the high degree of control over the various parameters and the purity of the motion signal, i.e. the ability to remove non-motion cues to the objects 3-D shape. We have previously developed a SFM stimulus that represents the orthographic projection of a transparent, rotating cylinder (Husain, Treue & Andersen, 1989; Treue, Husain & Andersen, 1991). Even with only a small number of dots defining the cylinder subjects report a vivid impression of an object or surface in depth rather than a group of points suggesting a process of spatial integration. Furthermore we were able to show a high degree of temporal integration because subjects showed an extended temporal buildup (up to over 1 sec) of the SFM percept even with displays in which individual dots were only present for short periods of time ($\sim$100 msec) before being replotted somewhere else on the cylinder.

*Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139, U.S.A.
†Division of Neuroscience, Baylor College of Medicine, Houston, Tex., U.S.A.
‡Division of Biology, California Institute of Technology, Pasadena, CA 91125, U.S.A.
§ATR Human Information Processing Laboratories, Japan.
‖Department of Computer Science, Wellesley College, Wellesley, MA 02181, U.S.A. [*Fax* 1 617 283 3642].
¶To whom all correspondence should be addressed.

This led us to propose that the visual system is able to integrate the depth information derived from the individual stimulus features spatially through a process of surface interpolation and that such an interpolation mechanism facilitates the integration of motion information over time (Husain *et al.*, 1989).

We use the term interpolation informally to describe a process in which the object depth values extracted at the stimulus features are used to reconstruct a complete surface in depth which fills-in depth values between the known depth values. The companion paper (Hildreth, Ando, Andersen & Treue, 1995) provides a discussion of possible implementations of such a process (see also Ando, 1992; Hildreth, Ando, Andersen & Treue, 1991). Recently, Saidpour, Braunstein and Hoffman (1992) also described interpolation in the perception of SFM.

SFM without a surface interpolation process would use the motion of stimulus features to extract the corresponding depth values only, leading to a "wire-frame", skeleton-like representation of the object that makes no implicit assumptions about depth values between features. This absence of an object representation beyond the visible features would make the perception of SFM based on displays with limited point lifetimes very difficult. But in fact we are able to perceive salient SFM with such displays (Dosher, Landy & Sperling, 1989; Husain *et al.*, 1989; Treue *et al.*, 1991).

In the experiments presented here we investigate the proposed surface interpolation process psychophysically and test several predictions which can be derived from our surface interpolation hypothesis. We further show that a number of previous observations regarding the extraction of SFM can be accounted for by our surface interpolation hypothesis. In the companion paper we demonstrate how to incorporate both surface interpolation and temporal integration into computational algorithms for the recovery of 3-D SFM.

The most prominent characteristic of any spatial interpolation process is that it "fills-in" featureless areas. In the first experiment we test the ability of human observers to detect such featureless stimulus areas as a function of their size. We show that as predicted by a surface interpolation process, subjects have great difficulties in detecting the presence of these featureless areas on the object.

One of the advantages of using a surface interpolation process in the extraction of SFM is that the extracted surface incorporates the information derived from individual features and preserves that information even after the disappearance of any individual feature. This allows features that appear separated in time to contribute to the recovery of the same object. The wire-frame approach described above only represents the depth values of the currently present features and thus shows no such temporal integration. The main purpose of the extracted surface could thus be to link features over time without them actually being used in the neural representation of the object. Alternatively, if the information from individual stimulus features is only used to allow for the extraction of the interpolated surface but is not preserved explicitly beyond the interpolation stage then the internal representation of the observed object is a surface rather than a collection of individual elements. By putting the behavior of the individual features and that of the object's surface in conflict we are able to demonstrate in the second experiment that the 3-D percept indeed follows the behavior of the surface rather than that of the individual features, in agreement with a rather fundamental role for the interpolated surface.

Given the richness of the visual world outside the psychophysical display a surface interpolation process used to recover object shapes should incorporate knowledge about object boundaries to spatially limit the interpolation process. In a final series of demonstrations we relate various observations by Ramachandran, Cobb and Rogers-Ramachandran (1988) to a SFM process that uses surface interpolation and boundary constraints.

Finally, any process of surface interpolation that has as its input two depth planes that coincide spatially in the 2-D input (as is the case for our transparent objects), has to perform some form of segmentation on the input to be able to interpolate the different depth planes independently. We argue that recent findings concerning the processing of motion transparency in the visual cortex of awake behaving monkeys puts the process of segmenting the front and back surface of transparent objects as early as Area V1 of primate visual cortex.

## METHODS

### *General aspects*

The basic stimulus used in all of the experiments described here is a moving random-dot pattern representing the parallel projection of a rotating transparent cylinder covered with points [Fig. 1(A,B,C)]. This stimulus is generated on a PDP 11-73 computer and then presented to the subjects [Fig. 1(D)] on a CRT screen in a dimly lit room. The subjects sit without restraint and view the screen binocularly from a distance of 57 cm. All experiments described here used a two-alternative forced-choice procedure and the subjects held a box with two buttons to record their responses.

An important characteristic of our stimulus is the use of limited point lifetimes. All the dots are present at specific positions on the cylinder only for a predetermined number of frames and are then randomly repositioned. Thus the projected image consists of individual points moving only through short trajectories. This allows us to limit the amount of information an individual feature can contribute to the recovery of the perceived 3-D shape of the object. It also enables us to keep the 2-D density distribution constant across the display throughout the rotation of the cylinder (for a more in-depth discussion of our stimulus and its generation see Treue *et al.*, 1991).

The parallel projection of a rotating cylinder generates a 2-D flow field with a velocity distribution described by a half sinusoid [Fig. 1(E)] in which stimulus features in the middle of the display move at a high velocity that drops to zero at the edges of the display where the dots reverse their direction and move along the opposite surface. Note

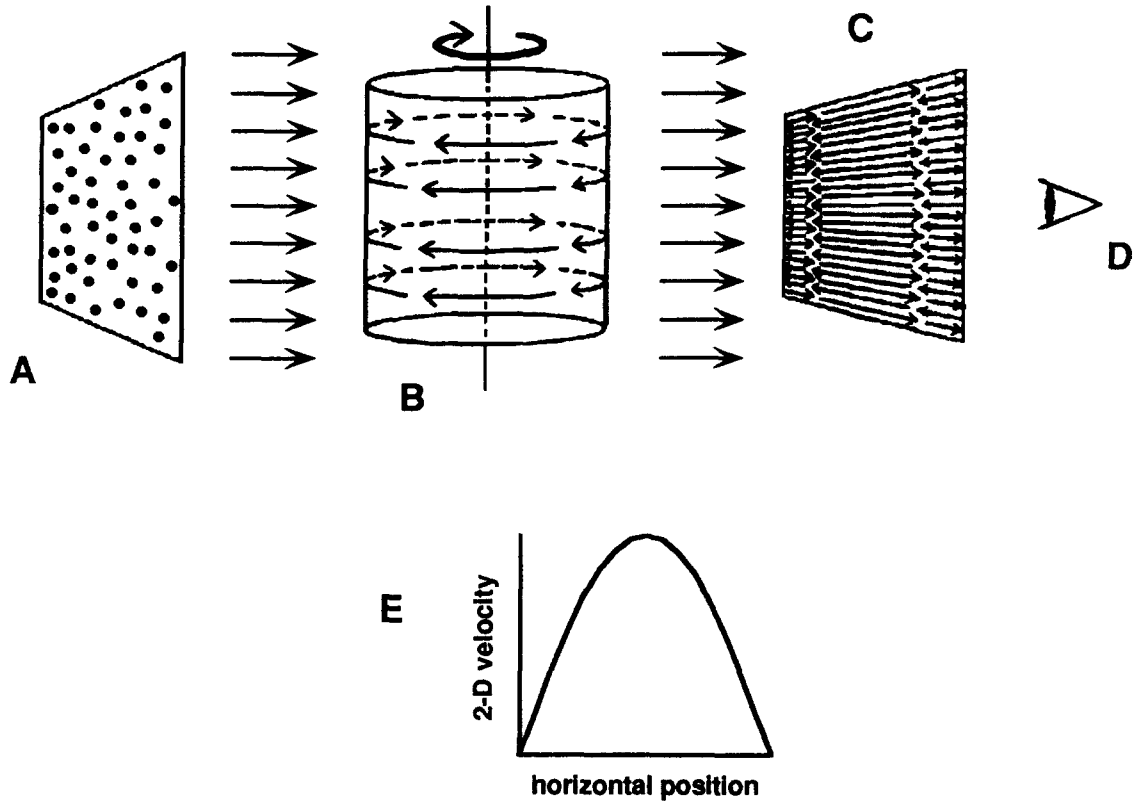


FIGURE 1. Cartoon of our stimulus generation algorithm. Points are plotted randomly on a 2-D square (A). They are then projected onto a cylinder (B) which is rotated. The projection of the rotating cylinder forms the visual display (C) that the subject observes (D). The velocity varies as a half-sinusoid across the display with the highest speeds in the center.

that the perceived direction of rotation for parallel projected objects is ambiguous and can reverse spontaneously during observation. This property is relevant for our Experiment 2.

## EXPERIMENTS

### *Experiment* 1

The use of a surface interpolation mechanism would allow the visual system to integrate information from points being present in the display at different times, and would also allow the visual system to reconstruct complete objects, i.e. surfaces from sparse data.

As mentioned in the Introduction one would expect a surface interpolation mechanism to fill in a surface in depth between the points in our stimulus. If the interpolated surface, rather than the individual dots form the internal representation of the object it would be difficult for the visual system to detect areas without features since the features are not explicitely represented. We set out to demonstrate this side-effect of interpolation by measuring how well subjects were able to detect the presence of masked parts or "holes" in the surface of a rotating cylinder as a function of the size of the occluded area. Normally such occluded areas would be easily detecteable by simple luminance differences to un-occluded areas but our transparent stimulus allows us to generate occluded parts without changing the local luminance in the image.

*Methods.* Subjects were presented with either a complete rotating cylinder or a rotating cylinder with a cut-out part. We investigated two mask conditions: in the first we masked one of four possible regions on either of the two surfaces of a rotating cylinder. These masked areas were stationary and centered on the middle of the four quadrants of the stimulus as sketched in Fig. 2(A). In the second condition the masked area was stationary on the surface of the cylinder and thus moved across the display during the stimulus presentation. The mask was randomly placed within a part of the cylinder that would not rotate from the front to the back or vice versa during the stimulus duration.

In both conditions the mask covered only one of the two surfaces, because otherwise the detection of a mask would amount to nothing more than a texture or luminance task, detecting stimulus patches (rather than patches on one surface only) void of points. For the same reason, the distribution of points was varied so that the surface patch opposite to the mask contained twice as many points as usual, thus guaranteeing a constant dot density across the stimulus. This density control could be easily implemented without artefacts because of our use of limited lifetimes. The projected size of the hole present in the second mask condition was kept constant while rotating around the cylinder so as to allow a better comparison to the first mask condition.

A further control was necessary in the first mask condition. Points on the surface of the cylinder disappeared when they rotated into or behind the

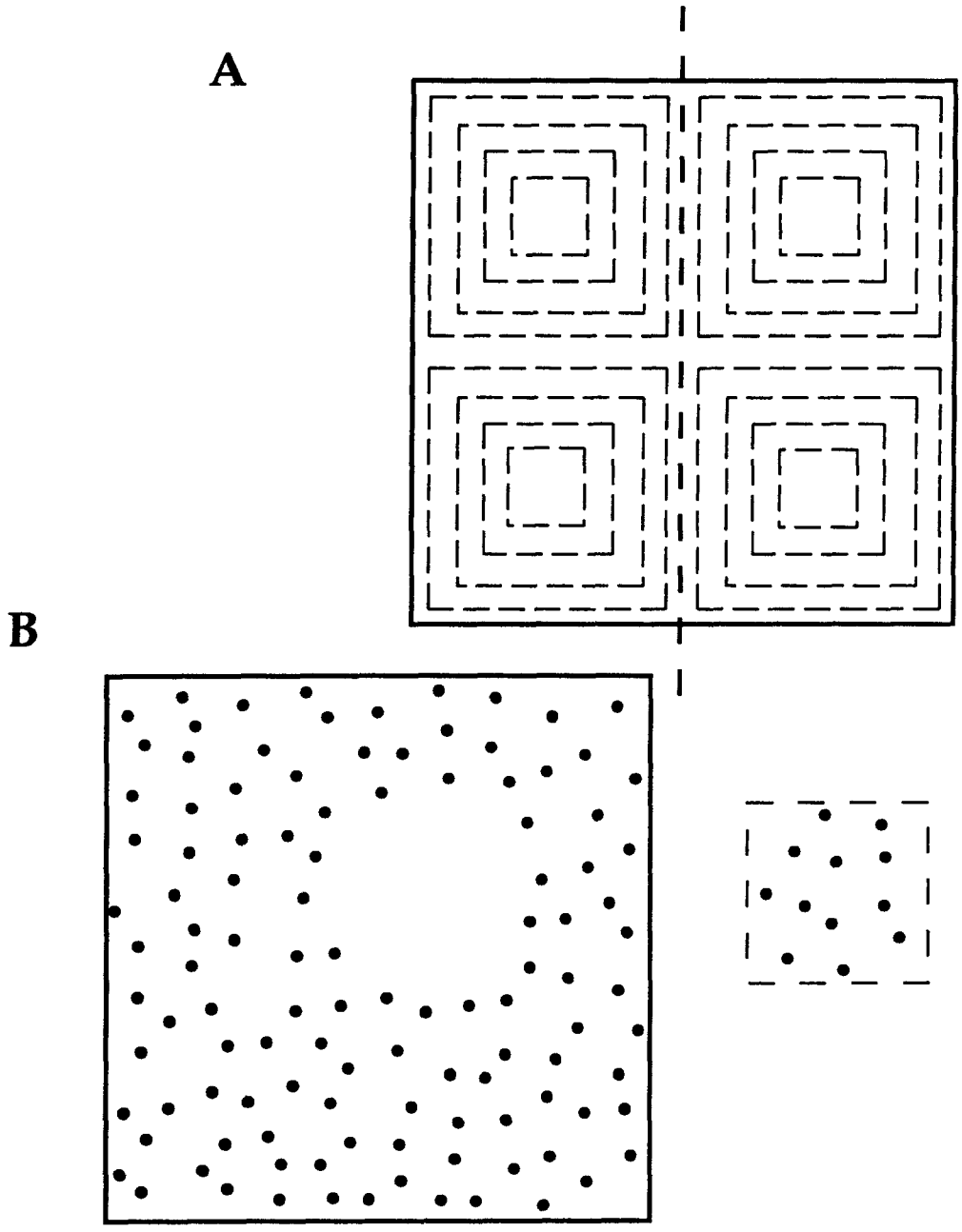


FIGURE 2. Mask positions for Expt 1. (A) To-scale drawing of the various mask sizes and positions used in Expt 1. (B) Example of the effect of using a mask. The cut-out stimulus part is represented on the right. This mask size represented the threshold of performance in Expt 1.

stationary mask in this condition and reappeared on the opposite side of the mask. This could serve as a cue to the subjects about the presence and location of the mask. To invalidate this spurious cue we positioned "virtual" masks in all four quadrants of the masked as well as of the unmasked cylinders. These virtual masks behaved like real masks in that points crossing their boundaries were replaced to the opposite side of these masks but they differed from real masks in that they contained moving dots rather than masking them. Through this manipulation virtual lines generated by disappearing and reappearing dots were present in all stimuli and could not serve as cues to the presence of a featureless area.

Subjects were presented with the stimuli which lasted 2 sec and rotated at an angular rotation rate of 50 deg/sec. They were instructed to press one of two buttons after the end of the stimulus to indicate whether they perceived the stimulus as a complete rotating cylinder or if they detected a mask or hole. The stimulus was made up of 125 points on each surface. The individual point lifetime was 200 msec, long enough for a strong impression of SFM (see Treue *et al.*, 1991) and short enough to introduce point disappearances and appearances that masked the edges of both the real mask as well as the "virtual" masks. These masks were squares of 2.25–20.25 angular deg$^2$ (the stimuli subtended $10 \times 10$ angular deg). The number of points masked thus ranged from about 3 to over 25 points. As an example Fig. 2(B) shows the effect of a mask of 3.5 deg width on one surface of the cylinder.

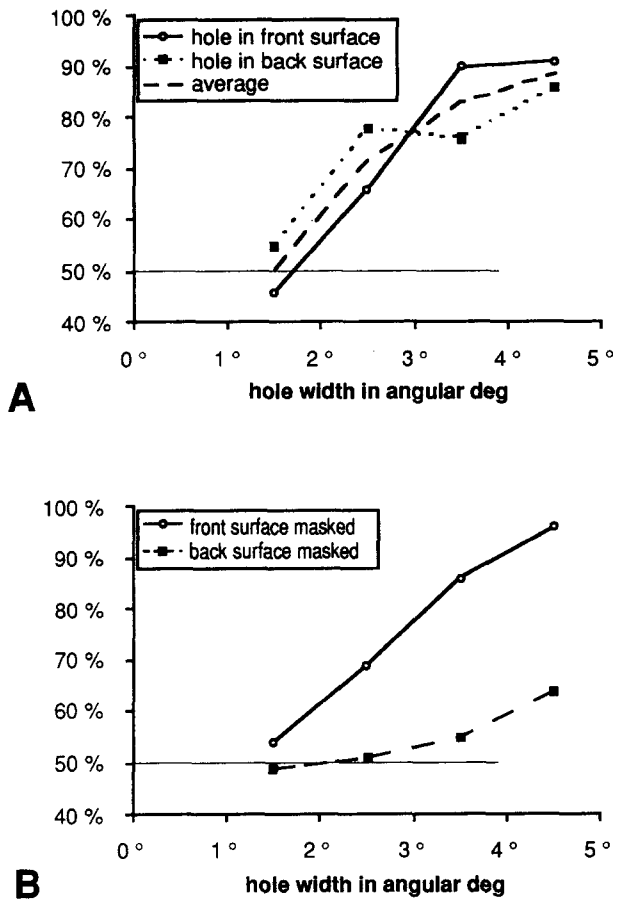


FIGURE 3. Experiment 1. (A) Performance detecting a hole in a cylinder as a function of hole size and hole position. (B) Performance detecting a masked part. Horizontal line represents chance level of performance. All values are averages across four subjects.

After a block of trials that typically consisted of 80 stimulus presentations, subjects were also asked which direction they perceived the cylinder as rotating.* With this information it was possible to plot performance separately for holes in the perceived front or perceived back of the cylinder.

*Results.* Figure 3 shows the results of the two experiments we performed. Figure 3(A) plots subjects' performance when presented with cylinders containing holes that moved with the rotation. Figure 3(B) plots the results of using masks that were stationary in 2-D. In both cases hole width in angular degrees is plotted on the $x$-axis while the two curves represent the performance for the hole being either on the perceived front or perceived back of the cylinder respectively. Chance in our two-alternative forced-choice paradigm was 50% and is denoted by a horizontal line.

Two results are apparent from the data. To be able to detect the presence of a hole or mask at the 83% level the size of the mask has to be about 10 deg$^2$, i.e. it has to cover nearly half of the stimulus quadrant it is placed in. In Fig. 2(B) we sketched the effect of a mask of that size on the 125 points present in one surface of the cylinder. This indicates that although subjects are able to easily segment the front and back in our transparent stimulus they have great difficulty evaluating the completeness of each surface. This result supports our hypothesis that the final percept of 3-D shape is more closely related to an interpolated surface representation than to the 3-D structure of individual features (see Expt 2).

Furthermore for the stationary mask there was a curious dependence of performance on the surface the mask was perceived to cover. Subjects were very poor at detecting even very large masks when they covered the perceived back surface of the cylinder. This difference between the two surfaces was not present for the hole that moved with the cylinder. Perceptually this seems to reflect the fact that a stationary "hole" in the back of the cylinder can only be achieved by the presence of a stationary non-transparent object within the cylinder. Given that subjects do not see such an occluder in our experiment they tend to ignore the possibility of a stationary mask covering the back surface.

*Experiment 2*

After strengthening the argument for the presence of surface interpolation in SFM we address the question of how fundamental a role this surface interpolation plays in the mental representation of the extracted object. Two possibilities come to mind. The surface interpolation might simply be used to allow the visual system to recover the 3-D positions of individual stimulus elements even when they are temporally separated without playing a role in the internal representation of the observed object. In such a scheme the interpolated surface would help in placing newly appearing dots in depth but would not play a role in the final representation of the stimulus, which would be as a group of points or a wire-frame style object positioned in 3-D. On the other hand, it is possible that the information from the individual features is only used to interpolate the surface, and that the object is ultimately represented through its surface rather than as a collection of individual elements in space or as a wire-frame style representation. If such a scheme were indeed employed by the visual system then the extraction of SFM should be determined by the behavior of the object's surface rather than by the behavior of the individual points.

Our stimulus allows us to perform an interesting variation to distinguish between these two schemes. Because we use a rotationally symmetric object in orthographic (parallel) projection the assignment of the front and back surface is arbitrary and in fact sometimes reverses spontaneously during viewing (similar to the Necker cube). Unlike the Necker cube this switch is not accompanied by a change in the object's surface shape or position, but rather only by a change in direction of apparent rotation.† Thus the perceptual reversal of a

*Fortunately subjects generally have such a strong bias for one direction of rotation that during our short stimulus durations they never perceived reversals of the perceived direction of rotation and also generally did not perceive different directions of rotations for the 80 trials in a block.

†Note that this is strictly true only if the patterns that define the front and back surface and generate the motion signal are indistinguishable. Otherwise patterns or other surface features have to move through depth from the front to the back surface and vice versa whenever the rotation direction changes.

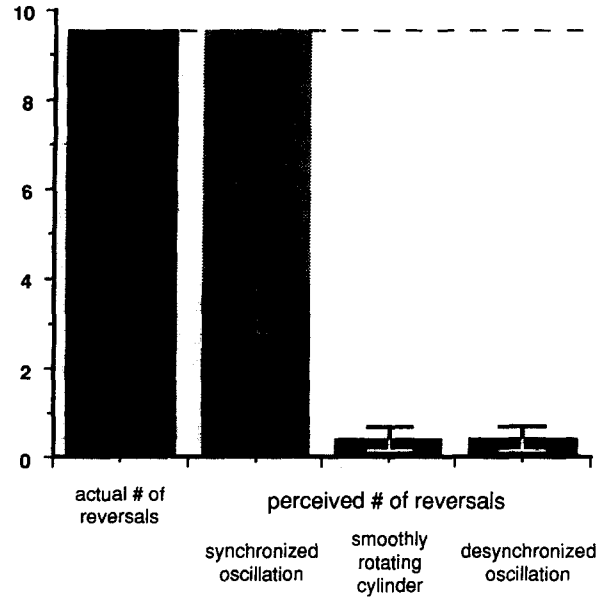


FIGURE 4. Experiment 2. First column: actual number of reversals in direction of motion. Second column: perceived number of reversals of motion (shaded bar represents actual reversals that the subjects failed to report). Third column: perceived number of spontaneous reversals of smoothly rotating cylinder. Fourth column: perceived number of reversals in stimulus made up of individual dots in asynchronous oscillations. Columns 2–4 represent the average of five subjects. The error bars reflect SEM.

Necker cube represents a physically unlikely event while the perceptual reversal of our rotating cylinder corresponds to a change in direction of rotation. We were thus interested in comparing these apparent changes in rotation direction with the perception of "real" (physical) changes in direction of rotation.

We presented subjects with a rotating cylinder in which all points reversed their direction of motion in synchrony after randomly chosen periods of time. The subjects' task was to indicate every perceived reversal of rotation direction by pressing a button. As a control we also measured the number of perceived reversals in a cylinder in which the individual points never reversed their direction of motion.

The results are plotted as the average of five subjects in Fig. 4. The first column indicates how often the points in the display changed their direction of motion during the stimulus duration. The second column is the perceived number of reversals of direction of rotation. The third column represents the control measure using a smoothly rotating cylinder. Two findings are immediately obvious. First the number of perceptual reversals in the smoothly rotating cylinder is rather low (in agreement with what the subjects reported in Expt 1). Secondly the subjects only saw about half of the "true" reversals as such. The shaded box on top of the second column represents the actual reversals that the subjects failed to respond to.

Two questions arise from this finding. How could the subjects miss so many of the reversals and how did they perceptually interpret the reversals that they did not see as such? When debriefing the subjects after the experiment, all reported two different percepts. Sometimes they saw the rotating cylinder reverse its direction (and pressed the button as instructed). At other times the cylinder seemed to stop momentarily and then continue to rotate in the same direction as before the stop. Neither of the two percepts was associated with any movements of individual dots in depth.

To account for this percept one has to assume that the visual system changed the assignment of all points from their current surface (front or back) to the opposite surface whenever they reversed their direction of motion. Only then can the cylinder as a whole be perceived to rotate in the same direction after the direction of motion of all the individual points is reversed. Thus the answer to the two questions posed above is: the subjects did see all instances of reversals of direction of motion of the individual points, but interpreted only some of them as a reversal of the rotation of the overall cylinder. Such a percept seems possible only if the object is internally represented as a 3-D surface rather than as a group of dots in space, because none of the subjects reported seeing the individual dots move through space from one surface to the other when the percept of "stopped" motion occurred. This is quite different from the perceptual reversals of the Necker cube, which is clearly interpreted as a change in 3-D location of the stimulus features.

It could be argued that the case presented to the subjects above is special in that a wire-frame representation of the cylinder could be maintained through the reversal of direction of motion since the 3-D distances between all points remain the same and all that happens is a depth reversal. One can think of the cylinder before and after the percept of stopped motion occurred as mirror images of each other. To control for that possibility we introduced a further modification to our stimulus.

We desynchronized the reversals of the individual points in time, i.e. between any two frames of our stimulus presentation only a small proportion of points reversed their direction of motion. If the visual system represents the points in the stimulus at their individually computed locations in space the stimulus would be interpreted as an entirely non-rigid cloud of dots since the two dots in any pair will sometimes move in the same and sometimes in the opposite direction. A wire frame representation of the stimulus would not just switch between two depth-reversed states but rather would be constantly changing since individual points would jump from the front to the back and vice versa in the process of changing their 3-D distance to all other elements of the object.

When we presented this stimulus to naive observers they all reported the percept of a smoothly rotating hollow cylinder. In fact we never found it possible to perceive anything but a smoothly rotating cylinder. When asked to count the number of reversals in direction of rotation subjects reported reversals as seldomly as they did for the uniformly rotating cylinder used as a control before (Fig. 4, two rightmost columns). To convince ourselves that the display indeed consisted of oscillating points we presented a stimulus in which one point was enlarged and was thus clearly distinguishable from the other features. When tracking this marked feature subjects reported it jumping through depth from the front

to the back surface or vice versa whenever it changed direction while the cylinder kept rotating in the same direction.

The percept of a smoothly rotating rigid cylinder in the face of the highly non-rigid physical stimulus can only be accounted for if the object is represented as a surface without an explicit representation of the individual features that contributed to it. On the other hand, if individual features are so distinct that they can easily be tracked over time the visual system seems able to preserve their individual positions in depth beyond the surface interpolation stage.

*Various observations and demonstrations*

In our previous studies (Husain *et al.*, 1989; Treue *et al.*, 1991) we investigated how precisely, how fast, and under which conditions humans can distinguish between a structured stimulus (the parallel projection of a rotating cylinder) and a control stimulus (generated by randomly shuffling the motion vectors of the structured display). But human perception of SFM is not always veridical and a variety of perceptual demonstrations (besides the one we documented above) have been documented in which subjects reported the percept of rotating cylinders or other objects when the display in fact was not physically consistent with such a percept.

Below we will describe several of these demonstrations as well as variations on them that we developed and discuss them in light of our proposed surface interpolation process.

*Segmenting multiple-surface displays*

Any surface interpolation process used to recover two surfaces present at the same image location (as is the case with our transparent cylinder) has to first segment the surface features based on the surface on which they lie before performing an interpolation. If a single surface were interpolated simultaneously through points laying on both the front and back of the cylinder, the result would be either a flat surface or a highly convoluted, continually changing shape, depending on the amount of smoothing performed. For rotating objects like the cylinder used in our experiments, such a segmentation process would be relatively easy since the two surfaces move in opposite directions.

Recordings from Area V1 in the awake behaving monkey suggest a physiologically plausible implementation of such a segmentation process. Snowden, Treue, Erickson and Andersen (1991) have recently demonstrated that transparent moving random-dot patterns activate two separate populations of direction tuned cells, one for each of the two opposite directions present in the display. On average, the response of the cells tuned to one direction is affected very little by the presence of points moving in the opposite direction. It thus seems that the visual system segments the two surfaces as early as V1 and this information might be used to guide the surface interpolation process. This line of reasoning is further supported by the results of our Expt 2. Given the "blindness" of direction-selective V1 cells for the presence or absence of their anti-preferred direction they would not be able to distinguish between our smoothly rotating cylinder and the two stimulus variations employed in Expt 2. These cells would only "see" one surface of the cylinder and points that belong to that surface and then reverse their direction of motion would simply "disappear". Such a point will then enter the group of points stimulating cells tuned for the opposite direction.

A visual illusion reported by Ramachandran *et al.* (1988) might be interpreted as evidence for a segmentation process more powerful than one based on opposite directions as described above. These researchers generated two displays representing two coaxial transparent rotating cylinders. In one display one cylinder had a smaller radius and its rotation speed was increased so that the 2-D velocity in the middle of the display was the same for both cylinders. Observers report a percept in which the two cylinders' surfaces seem to coincide in the middle and are separated in depth towards the edges of the display. In a variation of this display the two cylinders were of the same radius but one was rotated at twice the rotation rate as the other. Rather than perceiving the two cylinders as coinciding in depth, subjects reported a separation of their surfaces such that the faster rotating cylinder seemed to bulge out more in depth (see the companion paper for a more detailed description, analysis and figures). Since Ramachandran *et al.* used infinite point lifetimes and distributed points randomly on the cylinder (i.e. in 3-D) rather than on the display, these demonstrations contain an uneven density of points as a confounding depth cue.

We were able to replicate the results of Ramachandran *et al.* even after removing density gradients in the display by plotting the points randomly in 2-D and by the use of limited point lifetimes. These illusions might be interpreted as suggesting that the visual system can segment surfaces not only by opposite directions but also by their local 2-D speed. Such an inference might be premature, however, because we demonstrate in the companion paper that an algorithm that uses relative motion of features to determine their relative depth will generate comparable results even without surface interpolation and thus without any segmentation at all. The perceived segmentation of surfaces in these demonstrations might therefore not be due to an early segmentation process based on speed differences but might simply reflect the workings of a SFM process that assigns depth based on relative speed between stimulus features.

In relation to the multi-surface percept evoked by these stimuli it should be noted that subjects report difficulties perceiving all four surfaces (front and back of the two coaxial cylinders) at the same time. Rather they segment the attended surface (front or back) while perceiving a much less clear segmentation of the respective other surface. This is similar to findings by Andersen (1989) who reports that subjects can only detect up to three superimposed transparent surfaces moving in depth at a time. Because subjects' fixation was not monitored in any of these studies it is possible that maximally three surfaces

were perceived, since visual tracking could result in the image of one random-dot pattern being stationary on the retina and two patterns moving in opposite directions (thus again opening the possibility that the segmentation even in this case is based on opposite directions of motion).

*Effects of boundaries on SFM interpretation (Expt 3)*

Several studies have described displays in which boundaries influence the 3-D interpretation of moving random-dot patterns. Ramachandran *et al.* (1988) used two superimposed random-dot patterns moving with constant speed in opposite directions. When points reach the edge of the display they reverse their direction (they "bounce" off the edges). Ramachandran *et al.* report that subjects perceive the display as a rotating cylinder, rather than two flat planes. In a related demonstration they mask the projection of a rotating cylinder so that only a triangular or rectangular section is visible. They report that subjects describe a percept of a complete rotating cone or cylinder, respectively, rather than that of a masked, incomplete cylinder.

Thompson, Kersten and Knecht (1992) also report that under certain conditions a rectangular patch of random dots moving in one direction at constant speed surrounded by random-dot patterns moving at a different direction or at different speeds can result in the percept of a rotating cylinder. Royden, Baker and Allman (1988) report a similar finding for random-dot patterns moving within a rectangular patch surrounded by a stationary random-dot field. Craton and Yonas (1990) report that the spread of surface interpolation in motion perception is stopped at apparent stimulus boundaries.

Nakayama *et al.* (Nakayama, Shimojo & Silverman, 1989; Shimojo, Silverman & Nakayama, 1989; Nakayama & Shimojo, 1990) have recently suggested a framework that could help in interpreting these perceptual demonstrations. To capture the relationship between a perceived boundary and its two abudding surfaces they introduced the terms intrinsic and extrinsic. A boundary is intrinsic to a given surface if it is physically connected to the surface. In all the demonstrations described above that generated percepts of rotating cylinders, the boundaries were perceptually intrinsic to the cylinders. In most cases where occluders are used the boundary between the occluder and the object is perceptually intrinsic to the surface if the presence of the occluder is not recognized [Fig. 5(D)] while the border is perceptually extrinsic if the occluder is visible, e.g. as a cut-out floating in front of the object where the boundary is the edge of the cut-out [Fig. 5(C)].

We set out to test the hypothesis that the interpretation of a boundary as being intrinsic or extrinsic to a surface influences the role that the boundary plays in the reconstruction of the 3-D shape of the surface. Experiment 3 uses a display based on Ramachandran's observation that masking the sides of a vertically rotating cylinder results in the percept of a cylinder of smaller diameter, i.e. higher curvature. Our display contained four moving random-dot patterns. Figure 5 is a single frame out of the sequence of frames displayed on a computer screen. Figure 5(A) is the parallel projection of a transparent rotating cylinder very similar to the display we used for Expts 1 and 2 (except that we did not use limited point lifetimes here). Figure 5(C) is the same cylinder partially covered by a dark mask. Figure 5(D) represents a display similar to the one in Fig. 5(C) except that the mask is invisible. Figure 5(B) finally represents the parallel projection of a cylinder of a width equal to the width of the random-dot patterns in Fig. 5(C,D).

When subjects describe their percepts of display C and D they report that C seems to be part of a masked cylinder similar to the one in A. Random-dot pattern D, although physically identical to pattern C, is perceived as a cylinder of smaller diameter more like B (although generally not as highly curved as in B).

As described by Nakayama and colleagues (Nakayama *et al.*, 1989; Shimojo *et al.*, 1989; Nakayama & Shimojo, 1990) for the occlusion of surfaces perceived in depth, the visual system seems to distinguish between intrinsic and extrinsic surfaces in the recovery of SFM. The companion paper (Hildreth *et al.*, 1995) as well as Ando (1992) describes ways in which boundaries can interact with a surface interpolation mechanism to account for our observations. Here an intuitive explanation should suffice: extrinsic boundaries are ignored when the surface interpolation is performed since they are not part of the object and thus contain no depth information about the object. Intrinsic surfaces can influence the shape of the extracted surface since they are assumed to be part of the object. They serve as anchor points/lines that fall on the zero depth plane, i.e. are at the same depth as the center of the object in our display. They can thus be used as local depth estimates just like the values recovered from the individual dots in our displays. Having the object bounded on its side by lines at zero depth will pull the estimated surface toward that point, thus joining the front and back surfaces of our masked cylinder even though none of the dots fall near the zero depth plane.

## GENERAL DISCUSSION

In the experiments presented here we strengthened the case for the involvement of a process of surface interpolation in the recovery of 3-D SFM in the human visual system. Our findings suggest that the role of such a process goes beyond being simply a means for recovering the depth of individual feature elements presented in temporal separation. Rather our Expts 1 and 2 suggest that the internal representation of the object in the visual system is an object described by its surface and not a cloud of individual features. Experiments 1 and 2 show that even severe manipulation of individual features (i.e. out-of-phase oscillation of dots, or even removal of groups of dots by occlusion) remain unnoticed as long as those changes do not affect the interpolated surface.

Such a representation would provide an easy way for integrating other cues for depth perception which are

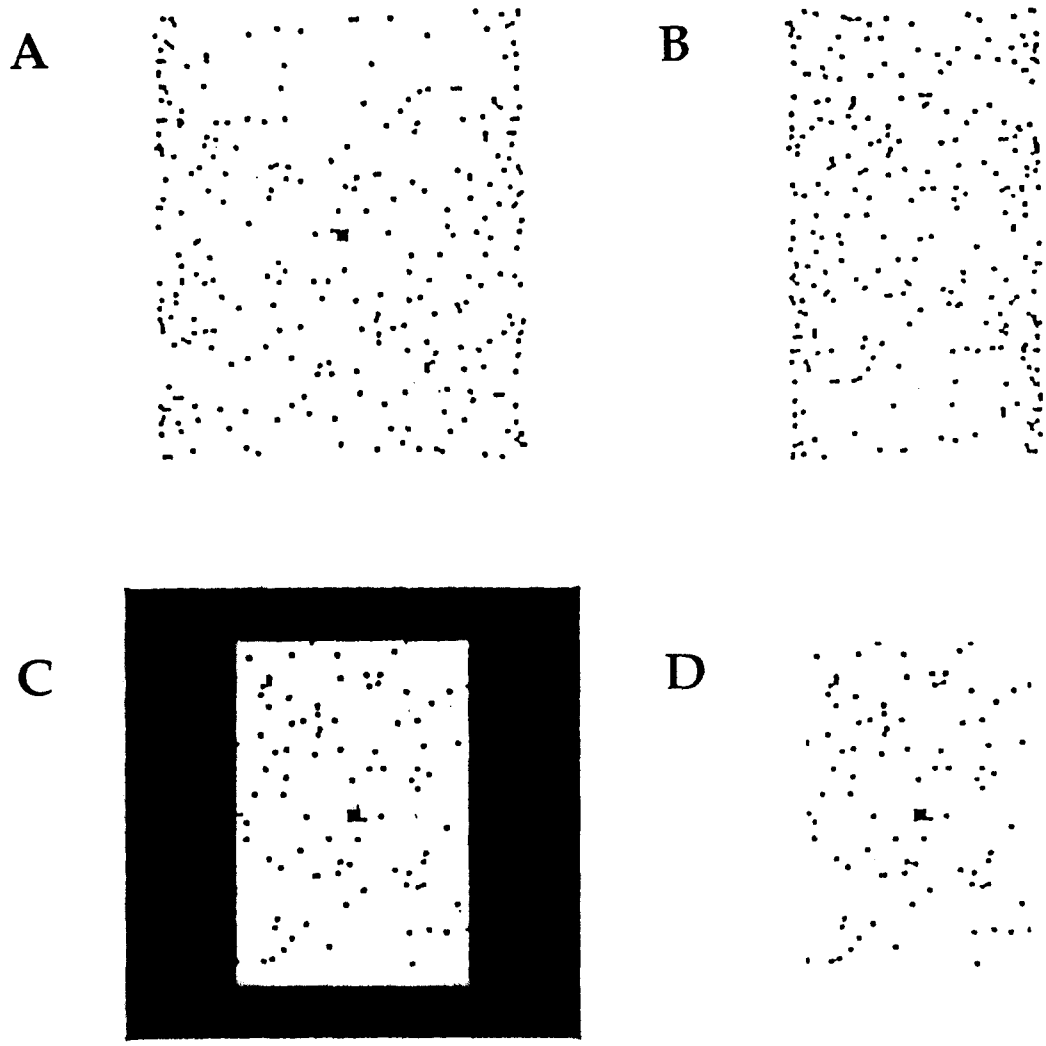


FIGURE 5. Individual frames from the display investigating the role of surface boundaries. (A) Complete cylinder. (B) Complete cylinder of smaller diameter. (C) Cylinder masked by visible mask (extrinsic border). (D) Cylinder masked by invisible mask (intrinsic border).

often surface based, like shape-from-shading, texture gradients and even binocular disparity, since there have been several reports for surface interpolation in stereoscopic depth perception (Collett, 1985; Mitchison & McKee, 1985; Mitchison & McKee, 1987; Buckley, Frisby & Mayhew, 1989; Würger & Landy, 1989).

This leaves open the question of how we perceive objects that, rather than having a distinct surface, represent a volume of points. Two explanations seem possible. Volumes could be represented in an onion skin like fashion as layers of surfaces. Alternatively, in the absence of distinct surfaces the visual system could resort to tracking individual groups of points to determine the axis of rotation as well as the range of 3-D distances from this axis present in the image. The possibility that the visual system could under certain circumstances extract depth from motion without interpolating a surface seems not unlikely given that no surfaces can be recovered in instances like Johansson's biological motion displays (Johansson, 1975). In these cases constraints derived from our knowledge of how humans move allow a very accurate representation of the 3-D motion from just a few strategically located moving points.

In the absence of such constraints Todd *et al.* (Todd & Bressan, 1990; Norman & Todd, 1993) using wire-frame objects in rotation to study affine stretching transformations find poor performance and little evidence for temporal integration. They show that subjects perform poorly on non-surface based tasks, as for example the estimation of 3-D line length, while they show good performance using similar stimuli when comparing the slant of surfaces formed by two intersecting lines in depth.

*Physiological implementation*

An important issue not addressed directly in the experiments reported here is the question of what kind of information the visual system uses as input to the depth recovery process. This is relevant in light of current computational approaches that use either the changes in the relative 2-D positions of object features (position-based approaches) or the velocity field of the projected object (velocity-based approaches). Previously we have

provided strong evidence that the visual system uses a velocity-based scheme (Treue *et al.*, 1991). Thus the algorithm presented in the companion paper uses such an approach. The specific implementation presented tracks the velocity of individual features. This is a computational convenience. It should be pointed out, however, that it is not easy to translate such a scheme directly into biological hardware, since the visual system with its stationary receptive field cannot really track individual features in dense random-dot patterns directly.* Rather the direction-selective neurons in the visual cortex act as spatio-temporal filters that generate a representation of the optical flow present in the retinal images. Through the aperture provided by the individual receptive fields the visual system is already performing a smoothing operation on the visual input since an individual neuron can only signal the overall motion in its receptive field and not the behavior of individual features. Although the visual system could still recover information about individual features through a careful combination of cells with overlapping receptive fields it is interesting to note that the overall activity in the population of neurons already represents a smoothed velocity field that is not keeping track of individual features *per se*, in agreement with what our Expts 1 and 2 suggest.

In summary, we have presented experiments and perceptual demonstrations that support the use of a surface interpolation scheme in the extraction of SFM in the human visual system. The companion paper presents a computational implementation of such a scheme that combines a feature-based SFM algorithm with a surface interpolation mechanism. The model allows multiple surfaces to be represented, incorporates constraints on surface structure from object boundaries, and segregates image features onto multiple surfaces on the basis of their 2-D image motion. The companion paper also presents the results of computer simulations that relate the behavior of the model to psychophysical observations.

## 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.

Ando, H. (1992). Dynamic reconstruction and integration of 3-D structure information. Ph.D., Department of Brain and Cognitive Sciences, MIT, Cambridge, Mass.

Buckley, D., Frisby, J. P. & Mayhew, J. E. (1989). Integration of stereo and texture cues in the formation of discontinuities during three-dimensional surface interpolation. *Perception, 18*, 563–588.

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

Craton, L. & Yonas, A. (1990). Kinetic occlusion: Further studies of the boundary flow cue. *Perception & Psychophysics, 47*, 169–179.

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.

Hildreth, E. C., Ando, H., Andersen, R. A. & Treue, S. (1991). Recovering three-dimensional structure from motion with surface reconstruction. AI Memo, 1314.

Hildreth, E. C., Ando, H., Andersen, R. A. & Treue, S. (1995). Recovering three-dimensional structure from motion with surface reconstruction. *Vision Research, 35*, 117–137.

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.

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

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

Mitchison, G. J. & McKee, S. P. (1987). Interpolation and the detection of fine structure in stereoscopic matching. *Vision Research, 27*, 295–302.

Nakayama, K. & Shimojo, S. (1990). Towards a neural understanding of visual surface representation. *Cold Spring Harbor Symposium on Quantitative Biology, 55*, 911–924.

Nakayama, K. & Tyler C. W. (1981). Psychophysical isolation of movement sensitivity by removal of familiar position cues. *Vision Research, 21*, 427–433.

Nakayama, K., Shimojo, S. & Silverman, G. H. (1989). Stereoscopic depth: Its relation to image segmentation, grouping and the recognition of occluded objects. *Perception, 18*, 55–68.

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

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

Royden, C., Baker, J. & Allman, J. (1988). Perception of depth elicited by occluded and shearing motions of random dots. *Perception, 17*, 289–296.

Saidpour, A., Braunstein, M. L. & Hoffman, D. D. (1992). Interpolation in structure from motion. *Perception & Psychophysics, 51*, 105–117.

Shimojo, S., Silverman, G. H. & Nakayama, K. (1989). Occlusion and the solution to the aperture problem for motion. *Vision Research, 29*, 619–626.

Snowden, R. J., Treue, S., Erickson, R. E. & Andersen, R. A. (1991). The response of area MT and V1 neurons to transparent motion. *Journal of Neuroscience, 11*, 2768–2785.

Thompson, W. B., Kersten, D. & Knecht, W. R. (1992). Structure-from-motion based on information at surface boundaries. *Biological Cybernetics, 66*, 327–333.

Todd, J. T. & Bressan, P. (1990). The perception of 3-dimensional affine stucture from minimal apparent motion sequences. *Perception & Psychophysics, 48*, 419–430.

Treue, S., Husain, M. & Andersen, R. (1991). Human perception of structure from motion. *Vision Research, 31*, 59–75.

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

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


*In fact that is the original motivation for Nakayama and Tyler's (1981) introduction of moving random-dot patterns into studies of the visual system.

*Acknowledgements*—This paper describes research done at the Center for Biological Information Processing, and Whitaker College at the Massachusetts Institute of Technology. The Centre's support is provided in part by the Office of Naval Research, Cognitive and Neural Sciences Division, the National Science Foundation (IRI-8719394 and IRI-8657824) and the McDonnell Foundation. This work was also supported by a grant to E. Hildreth and R. Andersen from the Educational Foundation of America and a grant to R. Andersen from the National Institutes of Health (EY 07492). S. Treue was supported by the Poitras Foundation.