Perception & Psychophysics
1988, 44 (4), 390-393

Notes and Comment

Perception of 3-D structure from motion: The role of velocity gradients and segmentation boundaries

V. S. RAMACHANDRAN, S. COBB,
and D. ROGERS-RAMACHANDRAN
University of California, San Diego, La Jolla, California

A topic that has received a great deal of attention recently is the problem of how we recover the three-dimensional (3-D) structure of moving objects. Consider the 2-D parallel projection of a rotating transparent 3-D cylinder with dots on its surface. The dots describe parallel horizontal paths, and the velocity of each dot varies sinusoidally as it moves from one side of the cylinder to the other and reverses direction. Although this changing pattern of dots is compatible with an infinite set of *non-rigid* interpretations (including that of a single plane of sinusoidally moving dots), observers always report seeing a rigid rotating 3-D cylinder, an effect that is often called the *kinetic depth effect* (Wallach & O'Connell, 1953) or *structure from motion* (Inada, Hildreth, Grzywacz, & Adelson, 1987; Schwartz & Sperling, 1983; Ullman, 1979).

One approach to this problem originated with Helmholtz (1925) and is based on the assumption that motion parallax and stereopsis are analogous, that is, that 3-D structure from motion is recovered from velocity gradients in much the same way that stereopsis is recovered from disparity gradients (Braunstein, 1962). For example, when an observer fixates any point in the world and moves sideways, then nearby objects appear to move in the opposite direction, whereas distant objects appear to move in the same direction as the observer. Furthermore, the velocities of the objects are proportional to their distances from the point of fixation, which implies that velocity gradients can potentially be used to determine relative depth (Braunstein, 1962). Ullman (1979) questioned the logical validity of this approach. Since different points in the visual world may actually be moving at different velocities, there is no a priori reason to assume a specific relationship between depth and velocity unless the points are connected together to constitute a rigid object. Ullman suggested that the derivation of 3-D structure from motion may be based, instead, on a special-purpose algorithm that seeks rigid interpretations. First, he showed mathematically that there is enough information in three views of four noncoplanar points to derive a unique 3-D structure if the assumption is made that the object is rigid. Second, he suggested that whenever the visual system is confronted with an ensemble of moving points, it applies a *rigidity test*; that is, it asks "does this collection of moving points have a unique interpretation as a rigid rotating object." If the answer is "yes," the system homes in on this rigid interpretation and discards the infinite set of nonrigid interpretations that are theoretically compatible with the same changing pattern.

The major strength of the computational approach to vision is that it allows a much more rigorous formulation of perceptual problems than would be possible with psychophysics or physiology alone. Ullman's (1979) elegant formulation of the structure-from-motion problem is a case in point. Before this formulation, there was a great deal of vague talk about how the visual system had a built-in "propensity" for seeing things rigid; it was assumed that we usually see things rigid because we *expect* them to be rigid. It was Ullman who first showed clearly that rigidity, far from being merely a vague propensity built into visual processing, powerfully constrains the solution to the structure-from-motion problem. In fact, his argument was that (1) without the rigidity assumption, the problem is unconstrainted and insoluble, and (2) if one assumes rigidity, then 3-D structure can be recovered without making any other assumptions.

One of our objectives has been to design experiments that might serve to distinguish between velocity-based and rigidity-based schemes for recovering 3-D structures from motion. Unfortunately, this has proved to be notoriously difficult in the past. For any rigid rotating object, the velocity of points in a parallel projection vary with their distance from the observer, and consequently, in most cases, both types of mechanisms yield the same solution. A critical test, however, would be to confront the observer with two coaxial cylinders of identical diameter spinning at different speeds (Ramachandran, 1985b). The velocity-based scheme would make the counterintuitive prediction that the faster cylinder should look more convex (since the velocity gradient is steeper), whereas the rigidity-based scheme would predict that the two cylinders should look identical. What do people actually see when confronted with such a display? We describe several demonstrations that were designed to answer this question and to investigate the mechanisms that the human visual system uses for recovering 3-D structure from motion.

Demonstration 1—Motion Parallax: Multiple Coaxial Cylinders

In this demonstration, we displayed two coaxial cylinders of identical diameter superimposed on each other and spinning at two different speeds, 5 rpm and 10 rpm. Can the visual system unscramble the two planes of dots

We thank Francis Crick and Dorothy Kleffner for helpful discussions. V. S. Ramachandran was supported by Biomedical and Academic Senate grants from the University of California. Address correspondence to V. S. Ramachandran, Department of Psychology, C-009, University of California, San Diego, La Jolla, CA 92093.

and perceive two rotating cylinders? Is the derivation of structure from motion possible under these conditions?

We found that in this display it was extremely difficult to perceive two cylinders of identical diameter spinning at different velocities (Ramachandran, 1985b; Ramachandran, Cobb, & Rogers-Ramachandran, 1987). Instead of seeing the dots occupy only the external surface of the cylinder, what was usually perceived was dots occupying two different depth planes and rotating with identical *angular* velocities, as though there were a small cylinder inside a large outer cylinder. The percept was a curious one since, on careful inspection, it was obvious that all the dots were in fact making identical horizontal excursions.

We believe that this illusion occurs because the brain has a strong propensity to translate velocity gradients into gradients of depth, as originally suggested by Helmholtz (1925). In fact, there may be a built-in assumption that if neighboring points have different velocities, then the slower ones must be nearer to the axis of rotation even though the ensemble of points has no rigid solution.[1] This interpretation is somewhat at odds with Ullman's (1979) contention that velocity gradients cannot directly specify gradients of depth or structure from motion. In this example, even though the visual system is given the opportunity for recovering a rigid solution, it actually rejects this interpretation and prefers to respond directly to velocity gradients.

Demonstration 2—Coaxial Cylinders of Dissimilar Diameters Rotating at Different Speeds

This display was similar to the previous one except that one of the cylinders was half the diameter of the other. Also, the smaller cylinder was made to spin at twice the angular (18 rpm) velocity of the larger one (9 rpm).

Structure from motion could easily be recovered from this display, and we usually saw two coaxial cylinders. As in the previous demonstration, however, the cylinders appeared to have the same angular velocity. When the dots on the smaller cylinder approached the middle of their excursion, they were seen to occupy almost the same depth as the dots on the outer cylinder. This is the converse of the result reported above, and it supports our contention that the visual system will assign identical depth values to dots that move at similar linear velocity even if they belong to different cylinders. Notice that this would require a nonrigid deformation of the cylinders, which implies that, as in the previous demonstration, the visual system will actually overcome rigidity in order to utilize velocity cues.

A formal experiment along these lines was conducted on 6 naive subjects who were unaware of the purpose of the experiment. They were shown two concentric coaxial cylinders of which the inner one was two-thirds the diameter of the outer one. (Their diameters were 2° and 3°, respectively.) The angular velocity of the outer cylinder was initially set by the experimenter at some randomly chosen value, and the subject's task was to vary the angular speed of the inner cylinder until its surface appeared to bulge forward and touch the outer one. We found that subjects could easily make this adjustment (Figure 1). They usually set the angular velocity of the smaller cylinder to be much higher than that of the larger one, confirming our initial impression that the visual system almost directly translates relative velocity into relative depth, even if this requires a considerable nonrigid deformation of the cylinders.

Demonstration 3—"Cylinder" of Dots Moving at Linear Instead of Sinusoidal Velocity

In this display, we had two transparent planes of dots superimposed on each other and moving in opposite directions at a constant linear velocity. As each dot reached one of the vertical borders, it simply reversed direction and retraced its path. Although this display is physically compatible with two flat coplanar sheets of dots moving in opposite directions, what we actually observed was a rotating 3-D cylinder. It was as though the mere reversal of direction at the border was sufficient basis for the brain to perceive a depth separation (and a curved motion path), even though the dots were actually moving at a constant linear velocity. The illusion was especially pronounced at high speeds of rotation (>30 rpm). Notice that no rigid interpretation is theoretically possible in this stimulus, yet the visual system recovers a 3-D shape that looks approximately rigid.

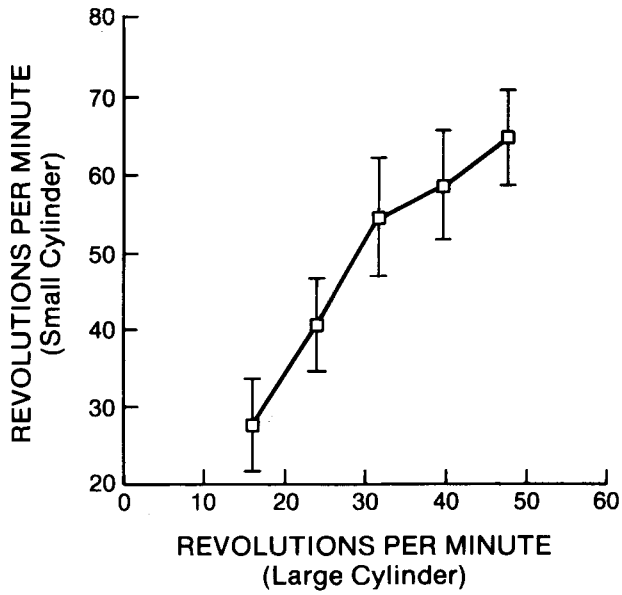

Figure 1. Data obtained from two concentric coaxial cylinders. The speed (rpm) of the inner cylinder was adjusted by the observer until its surface appeared to bulge forward to touch the surface of the outer cylinder. Each datum point represents the mean of 240 readings (6 subjects × 40 readings each). Vertical lines indicate standard deviation from the mean.

Demonstration 4—The Role of Segmentation Boundaries: Cylinder Viewed Through a Triangular Aperture

We began with a transparent 3-D cylinder and viewed it through a triangular "window" or aperture, so that only a triangular patch of moving dots was visible (the horizontal base of the triangle was exactly equal in width to the diameter of the cylinder). The display was viewed in complete darkness so that the occluder was not visible. To our astonishment, this display looked very much like a solid 3-D cone rather than part of a cylinder. Even though there was no velocity gradient along the vertical axis, the dots near the base of the cone were perceived as being further from the axis of rotation that the dots near the apex at the top. This observation implies that although velocity gradients are often sufficient to specify 3-D structure from motion, they are not necessary. Furthermore, the *segmentation boundaries* that delineate the object in motion (i.e., the edges of the triangular window) seem to have a strong influence on the magnitude of perceived depth.

Similar results were obtained when the cylinder was viewed through a vertical rectangular window whose width was about half that of the cylinder. This display looked like a much smaller cylinder, again suggesting that the magnitude of depth perceived for any given velocity gradient is strongly influenced by the total horizontal width of the object.

Demonstration 5—Two Adjacent Rocking Cylinders

Our last demonstration also suggested that the depth perceived in these displays was strongly influenced by image segmentation. We generated two adjacent transparent 3-D cylinders on the CRT, each of which was composed of 15 dots and subtended 2° (width) × 3° (height). Each cylinder was then made to "rock" (i.e., to make short clockwise and counterclockwise rotations) on its long axis, and this was sufficient to convey a strong impression of 3-D structure. We then moved the two cylinders horizontally toward each other until they almost touched, so that we ended up with one large cluster of dots rather than two separate clusters. We found that this display then looked like a *single* large cylinder rotating (rocking) on its axis rather than two cylinders. Dots near the middle of the large cluster appeared to move forward toward the observer much more than when either cluster was viewed separately. One could produce striking changes in the magnitude of perceived depth in the z-axis simply by alternately adding or deleting one of the two dot clusters (cylinders) that composed the single large cluster. The effect was especially striking if the display was viewed in complete darkness so that the margins of the CRT screen were not visible.

This result is interesting for two reasons. First, since the two cylinders do not share a common axis (i.e., since they have separate axes of rotation), the ensemble of points has no rigid solution. Nevertheless, the visual system tends to see the entire collection of points as constituting a single object that is approximately rigid. Second, the magnitude of depth perceived is much greater in this "fused" object than in either of the two separate cylinders, again suggesting that the total horizontal width of the object (as revealed by segmentation boundaries) can strongly influence the magnitude of perceived depth. We have previously demonstrated the important role played by segmentation boundaries for a variety of other perceptual capacities, such as the perception of apparent motion (Ramachandran, 1985a; Ramachandran & Anstis, 1986), stereopsis (Ramachandran, 1986), and shape from shading (Ramachandran, 1988).

Using Multiple Strategies for Recovering Structure from Motion

Ullman (1979) showed that 3-D structure from motion can, in principle, be recovered from the changing shadow if the assumption is made that the object producing the shadow is rigid. Given the rigidity assumption, Ullman's structure-from-motion theorem proves that there is enough information available in three views of four noncoplanar points to uniquely specify their 3-D structure. However, our results imply that the particular algorithms suggested by directly applying the methods used in mathematical proofs of the theorem are unlikely to be the ones actually used by the human visual system. Demonstrations 1 and 2 are two examples in which velocity cues and rigidity are pitted against each other, and in both situations velocity seems to win. In fact, the system seems to be quite willing to overcome rigidity in order to adhere to the velocity equals depth rule.

Our results suggest that the recovery of 3-D structure from motion may be analogous to the perception of shape from shading (Ramachandran, 1988), in that it relies on the *combined* use of velocity gradients and segmentation boundaries. It does not follow from this, however, that the system never uses a rigidity-based algorithm of the kind suggested by Ullman (1979). For example, consider the case in which a small number of widely separated dots spin (or rock) around a single axis. If the display is very large (e.g., subtends 20° or 30°), it strains the imagination to think of segmentation boundaries in the usual sense, and perhaps one would have to resort to the use of a rigidity-based scheme.

The velocity scheme leaves one important question unanswered: How does the visual system know what scaling factor is to be used in translating relative velocities into relative depth? Ullman (1983) showed that any given velocity gradient is, in fact, theoretically compatible with a whole family of surfaces (including nonrigid ones). How would the visual system know which one to pick? One possibility is that the system sets the scale simply by using the object's outline; that is, it may use the total horizontal width of the object to adjust the gain of the mechanism that translates velocity gradients into depth gradients. This would explain the critical role played by segmentation boundaries in Demonstrations 4 and 5.

Taken collectively, our findings imply that, in addition to Ullman's (1979) rigidity algorithm, the brain also appears to use a variety of other gimmicks to recover 3-D structure from motion. However, if the rigidity algorithm alone will suffice, theoretically why does the system resort to using so many other strategies? This question can be raised for all aspects of perception, of course, and not only for the structure-from-motion problem. Perhaps the simultaneous use of a wide range of shortcuts allows the visual system to tolerate "noisy" stimuli and to achieve more rapid processing of visual information than it could with a single sophisticated algorithm (Ramachandran, 1985c).

REFERENCES

BRAUNSTEIN, M. (1962). Depth perception in rotation dot patterns: Effects of numerosity and perspective. *Journal of Experimental Psychology*, **64**, 415-420.

HELMHOLTZ, H. VON (1925). *Treatise on physiological optics* (J. P. Southhall, Trans.). New York: Dover. (Original work published 1909)

INADA, V., HILDRETH, E., GRZYWACZ, N., & ADELSON, E. H. (1987). The perceptual build-up of 3-D structure from motion. *Investigative Ophthalmology & Visual Sciences*, **28**(Suppl.), 142.

RAMACHANDRAN, V. S. (1985a). Apparent motion of subjective surfaces. *Perception*, **14**, 127-134.

RAMACHANDRAN, V. S. (1985b). Inertia of moving visual textures. *Investigative Ophthalmology & Visual Sciences*, **26** (Suppl.), 56.

RAMACHANDRAN, V. S. (1985c). The neurobiology of perception [Guest editorial for special issue on human motion perception]. *Perception*, **14**, 1-7.

RAMACHANDRAN, V. S. (1986). Capture of stereopsis and apparent motion by illusory contours. *Perception & Psychophysics*, **39**, 361-373.

RAMACHANDRAN, V. S. (1988). Perception of shape from shading. *Nature*, **331**, 163-166.

RAMACHANDRAN, V. S. (in press). Interactions between motion, depth, color, form and texture: The utilitarian theory of perception. In C. Blakemore (Ed.), *Vision: Coding and efficiency*. Cambridge, England: Cambridge University Press.

RAMACHANDRAN, V. S., & ANSTIS, S. M. (1986). Perception of apparent motion. *Scientific American*, **254**, 102-110.

RAMACHANDRAN, V. S., COBB, S., & ROGERS-RAMACHANDRAN, D. (1987). Recovering 3-D structure from motion. *Society for Neurosciences Abstracts*, **13**, 630.

SCHWARTZ, B. J., & SPERLING, G. (1983). Nonrigid 3-D percepts from 2-D representations of rigid objects. *Investigative Ophthalmology & Visual Science*, **24**(Suppl.), 239.

ULLMAN, S. (1979). *The interpretation of visual motion*. Cambridge, MA: MIT Press.

ULLMAN, S. (1983). Recent computational studies in the interpretation of structure from motion. In J. Beck, B. Hope, & A. Rosenfield (Eds.), *Human and machine vision* (pp. 315-328). New York: Academic Press.

WALLACH, H., & O'CONNELL, D. N. (1953). The kinetic depth effect. *Journal of Experimental Psychology*, **45**, 205-217.

NOTE

1. This implies that for the front surface of the cylinder, the faster dots are nearer to the observer, whereas for the back surface, the faster dots are further away from the observer.

(Manuscript received June 26, 1987;
revision accepted for publication April 12, 1988.)