From *ACM Siggraph and Sigart Interdisciplinary Workshop on Motion: Representation and Perception* (pp. 11-16), Toronto, April 4-6 (1983).

THE PERCEPTION OF COHERENT MOTION IN TWO-DIMENSIONAL PATTERNS

Edward H. Adelson
RCA David Sarnoff Research Center

and

J. Anthony Movshon
New York University

Introduction

When one looks at a two-dimensional scene of moving objects, one can usually assign a velocity to each point in that scene with little effort. This suggests that some early visual processes are able to generate a two-dimensional velocity map using fast parallel computations. But it is not obvious how this should be done, and we are currently trying to understand how the human visual system does it.

If moving patterns were merely one-dimensional, the visual system's task would be easier. One dimensional motion can be signaled by cells such as those described by Barlow and Hill in rabbit (1963); these units compare responses of two adjacent regions on the retina at two successive moments of time. Combinations of such cells can signal the one-dimensional velocity of a moving edge, bar, or other contour.

But the retinal image is two-dimensional, and the problem becomes more complex than it first seems. Figure l illustrates a well-known visual phenomenon called the "barberpole effect," which is based on the inherent ambiguity of the motion of extended contours.

In fig. 1(a), the grating seems to move to the right when viewed through the horizontal window, even though the "true" motion of the physical grating is diagonally down and to the right. When a vertical window is used, as in fig. 1(b), the same physical motion leads to a percept of a downward motion.

The fact that one physical motion can be seen in different ways is to be expected because the observed motion of a stimulus like a grating does not, in and of itself, define the physical motion that produced it. The point is made clear in figure 2, where three different physical motions give rise to an identical pattern of stimulation when viewed through the window. In this case, all three patterns will be seen as moving down and to the right.

The ambiguity is not, of course, limited to the motion of gratings. Any extended pattern, such as an edge or a straight line, will offer the same problem. Consider the two moving diamonds shown in fig. 3. In fig. 3(a), the diamond moves down, yet a cell that "looks" at a local patch on the lower right edge will signal a contour moving down and to the right. In fig. 3(b), the diamond moves to the right, and the same cell will again signal contour motion down and to the right. So it is impossible to tell which way the diamond is moving by merely looking at one of its edges.

Marr and Ullman (1981) call this the "aperture problem," and have discussed how one can combine information from cells that signal only direction in order to constrain the underlying motion to lie within a limited range of directions. Fennema and Thompson (1979) showed that by taking advantage of local information about both the direction and speed of moving contours, one could obtain a full solution for the object's motion using a Hough transform; we take a similar approach in the work we will discuss. (Note that this solution is only correct for the case of pure translation of a two dimensional pattern. For more complex transformations, such as rotations or deformations, more complex strategies are required (Horn and Schunck, 1981; Hildreth and Ullman, 1982)). An early discussion of the problem can be found in Wohlgemuth’s (l9ll) classic work on the motion aftereffect; Hans Wallach (1935, 1976) did some very interesting experiments on its perceptual implications. It continues to interest both physiologists and psychologists today (Henry et al, 1974; Burt and Sperling, 1981; Adelson and Movshon, 1982).

There is both physiological and psychophysical evidence that the first stages of visual processing analyze the retinal image into a patchwork of localized one-dimensional components, which may variously be conceived as representing bars, edges, local Fourier components, Gabor functions, or what have you. In any event, such an analysis brings the aperture problem in with it from the very start. The visual system must go from the local motion of one-dimensional components to the percept of a single coherently moving pattern. In our discussion we will use the term component motion to refer to the motion of extended one-dimensional patterns such as lines, edges, and gratings; and we will use the term "pattern motion" to refer to the unambiguous two-dimensional motion of more complex patterns such as textures and objects.

Resolving Ambiguity

While it is true that a single component motion is ambiguous, the ambiguity is limited to a single degree of freedom. Consider once again the motion of the lower right edge of a rightward-moving diamond; the edge is magnified in fig. 4. The set of object motions consistent with the observed edge-motion are indicated by the arrows fanning out from the edge. This family of velocities may be depicted as a line in "velocity space," where any velocity is represented as a vector from the origin whose length is proportional to speed and whose angle corresponds to direction of motion. The family of motions consistent with the edge maps to a straight line in velocity space.

Figure 5 shows how a pair of such constraints can be used to determine the true motion of the diamond. Each of the edges in fig. 5(a) is associated with its own family of motions in velocity space, as shown by the lines in fig. 5(b). The lines intersect in a single point, and that point represents the object's motion.

The same analysis may be applied to combinations of gratings, as shown in fig. 6. Two gratings move behind a circular window. One moves up and to the right, the other moves down and to the right. When they are added together the resulting plaid moves rightward, as an apparently rigid pattern, in accord with the requirements of the velocity space construction. The appearance of this pattern is interesting, because one does not see either of the component grating motions in the combined pattern. Rather, the plaid seem to move rightward as a single coherent surface.

At first glance, it may appear that we have described a very complex method for doing vector addition. But the velocity space solution is generally different than one would get from a vector sum, a point that is exemplified in figure 7. Here we combine two gratings, each of which moves down and to the right, each with its own speed and direction. On a vector sum model, the combined pattern should move down and to the right as well, while on a velocity space model it should move up and to the right. And the stimulus does look like it is moving up and to the right, in accord with the velocity space requirements.

Crossed gratings do not always cohere into a single moving pattern. Figure 8 shows a case in which they cannot do so because there are three gratings which generate mutually incompatible constraints. The three gratings here set up three constraint lines; any two gratings can cohere, but then the third one cannot be included in the same pattern motion. When one views this stimulus, one sees a multistable display, in which any two of the gratings can be seen as a coherent pair moving in one direction, while the third grating (the odd man out) floats off by itself in the opposite direction. There are three such percepts, each corresponding to a particular intersection in velocity space. One can select out a particular percept by tracking the desired pattern motion with one's eyes; the tracking strongly biases the perception toward the tracked coherent pattern rather then the other two possibilities.

Determinants of Coherence

Even with just two gratings, coherence does not always occur. In some circumstances a pair of crossed gratings, each moving in a different direction, will be seen as just that -- a pair of crossed gratings, each moving in a different direction, each sliding across the other as if the other wasn't there. In such cases the visual system has elected not to combine the two component motions into a single pattern motion. By studying the conditions under which coherence does and does not occur we may learn something about the mechanisms that underly the perception of pattern motion.

One of the most striking determinants of coherence is the spatial frequency of the crossed gratings (Adelson and Movshon, 1982). In many of our experiments we use sine-wave gratings, i.e., gratings whose luminance profiles are sinusoidal (such stimuli have become popular in vision research because many early visual mechanisms are spatial frequency tuned, so that by using a sine-wave stimulus one can preferentially stimulate a relatively small subset of the mechanisms under study). We have found that two gratings will have a strong tendency to cohere if they are of the same spatial frequency, but have a rather weak tendency to cohere if their spatial frequencies differ by more than an octave. Thus, for example, if a 3 cycle/deg grating of one orientation is summed with a 9 cycle/deg grating of a different orientation, chances are that they will be seen as sliding over each other rather than moving as a single coherent plaid.

The contrasts of the two gratings are also important in determining coherence. If the first grating is of high contrast while the second one is of low contrast, then coherence may break down. Only when the contrast of the second grating is increased will the coherent percept return.

The contrast dependence can be used to derive a tuning curve for the frequency dependence, in the following way. Start with one moving grating of fixed spatial frequency and contrast. Add to it a second moving grating (of a different orientation), of a different spatial frequency. If the second grating’s spatial frequency is substantially different from that of the first grating, then its contrast will have to be quite high in order for the two gratings to cohere. But if the second grating has a spatial frequency that is similar or identical to that of the first, then it will cohere with the first even when its contrast is quite low. By measuring the minimum contrast at which coherence occurs, one can trace out a tuning curve for this effect.

Figure 9 shows the results of two such experiments on one subject. The closed circles show the first experiment, in which the standard grating was 2.2 cycle/ deg, moving at 3 deg/sec, and had a contrast of 0.3. The second grating had an orientation that differed from that of the first grating by 135 degrees; the second grating’s spatial frequency was varied. The filled leaf circles indicate the minimum contrast the second grating needed in order to cohere with the first, at various spatial frequencies. The contrast was noticeably elevated when the two frequencies differed by an octave, and became quite high when they differed by two octaves. The tendency to cohere was greatest (and the needed contrast was lowest) when the second grating's spatial frequency was matched to that of the first grating.

The open circles show the results of a similar experiment in which the standard grating had a spatial frequency of 1.2 cycle/deg. The results are much the same, and once again the tuning curve is centered on the spatial frequency of the standard grating. So these experiments indicate that the visual mechanisms underlying coherent motion perception are spatial frequency tuned, like many other aspects of early visual processing.

Global Versus Local Analyses

The velocity space construction is quite useful in understanding and predicting pattern motion phenomena, but this utility does not demonstrate that the human visual system is actually performing the rather global computations suggested in the velocity space diagrams. There are alternative approaches that will lead to the unambiguous perception of moving patterns, such as approaches based on "landmarks," or localized features in the moving patterns. In the case of the moving diamond, a corner could serve as a landmark, and by following its position over time one could correctly determine the motion of the diamond. Similarly, in the case of crossed sine-wave gratings, the peaks and troughs corresponding to the intersections of the light and dark grating bars could serve as landmarks by which the coherent motion direction could be inferred. The motion derived by these approaches is, of course, identical to that derived from velocity space, since in either case there exists a single pattern motion that is consistent with all the visual information in the display.

But there is an interesting display that we call the "split herringbone," for which a landmark model makes a different prediction than does a more global model based in velocity space. The display is shown in fig. 10(a); it consists of alternating columns of line segments that tilt left or right on the odd and even columns. The odd (right-tilting) columns move down, while the even (left-tilting) columns move up.

The most obvious landmarks in this display are the endpoints of the line segments. If they determine the motion percept, then one should see the split herringbone for what it is: a set of interleaved columns moving continuously up or down.

If, on the other hand, a more global process is at work, then it is possible that one will see an illusion of rightward motion, as illustrated in fig. 10(b). The odd columns produce one constraint line in velocity space, and the even ones produce another. The intersection corresponds to pure rightward motion. So this global approach predicts an illusion of motion to the right.

When one actually sets up the display, one finds that either of the two percepts is possible. The model based on local landmarks works when the display is of high contrast, is sharp, and is centrally fixated; in these conditions one sees the "correct" percept of vertically moving columns. On the other hand, when the display is of low contrast, or is blurred with a diffusion screen, or is viewed peripherally, then one can see the illusion of rightward motion predicted by the global model based in velocity space.

When the illusion does occur it is quite striking. The herringbone pattern seems to be moving continuously rightward, and yet it never gets anywhere, since the vertical "creases" where the columns abut are fixed in position.

These observations suggest that both kinds of models -- the local landmark-based models, and the more global models based in velocity space -- can be useful in understanding the way we perceive moving patterns.

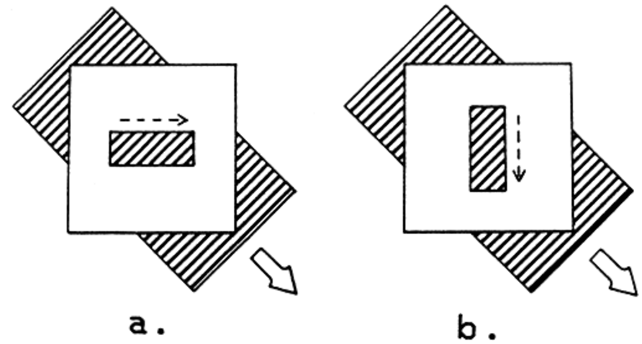

Fig. 1: The barberpole effect. The same physical motion of the grating behind the window can be seen as corresponding to different motions within the window.

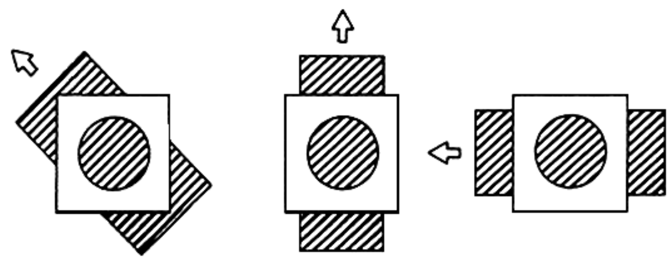

Fig. 2: The ambiguity of motion for extended contours. All three physical motions give rise to identical motion as viewed through the window.

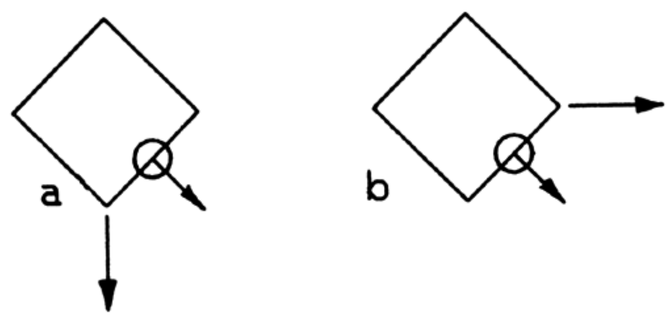

Fig. 3: A cell looking at a local edge of a diamond cannot determine which way the diamond is moving.

References

Adelson, E.H., and Movshon, J.A., Nature, 300, 523-525 (1982).

Barlow, H.B., and Hill, R.M., Science, 139, 412 (1963).

Burt, P., and Sperling, G., Psychological Review, 88, 171-195 (1981).

Henry, G.H., Bishop, P.O., and Dreher, B., Vision Research, 14, 767-777 (1974).

Hildreth, E.C., and Ullman, S., MIT AI Memo 699, December 1982.

Horn, B.K.P., and Schunck, B., Artificial Intelligence, 17, 185-203, (1981).

Fennema, C.L., and Thompson, W .B., Computer Graphics and Image Processing, 9, 301-315 (1979).

Marr, D., and Ullman, S., Proceedings of the Royal Society, London, B. 211, 151-180 (1981).

Wallach, H., Psychol. Forsch., 20, 325-380 (1935).

Wallach, H., On Perception, 201-216, Quadrangle, New York, 1976.

Wohlgemuth, A., British Journal of Psychology, monograph supplement 1(1911).

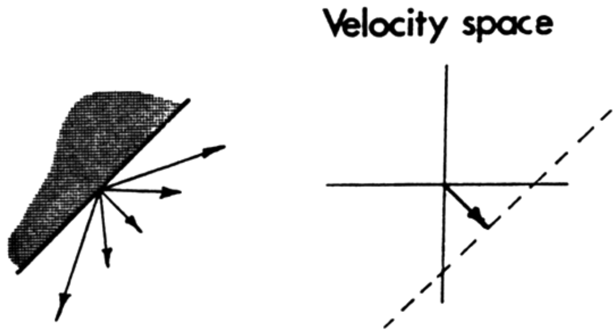

Fig. 4: Left: The set of possible object motions that could give rise to an observed motion of the edge. Right: Mapping the family of possible motions into velocity space.

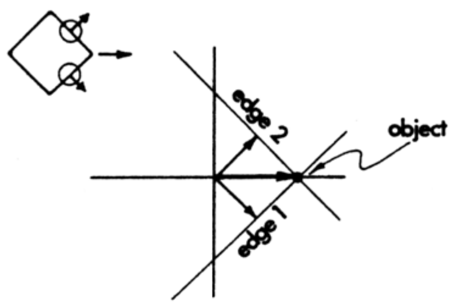

Fig. 5: An exact solution for the diamond’s motion can be determined by the intersection of two constraint lines from two edges.

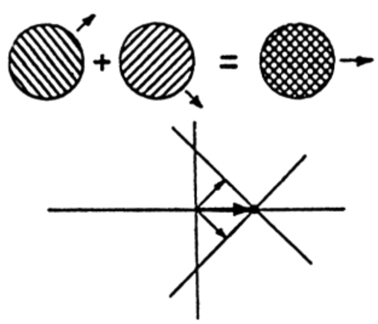

Fig. 6: The velocity space solution applied to the case of two moving gratings which are summed to form a moving plaid. The plaid moves to the right.

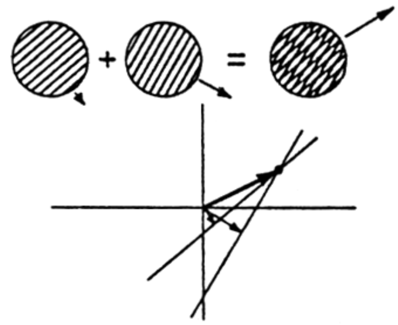

Fig. 7: An example where the velocity space solution is quite different from the solution based on a vector sum. Two gratings that move down and to the right give a pattern motion that is up and to the right.

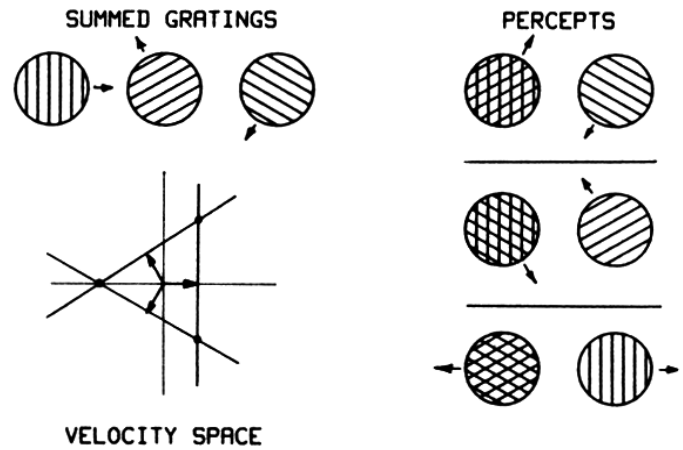

Fig. 8: A tristable display, composed of the sum of three moving gratings. There are three velocity space solutions, each of which is consistent with only two of the three grating motions. There are three possible percepts corresponding to these solutions.

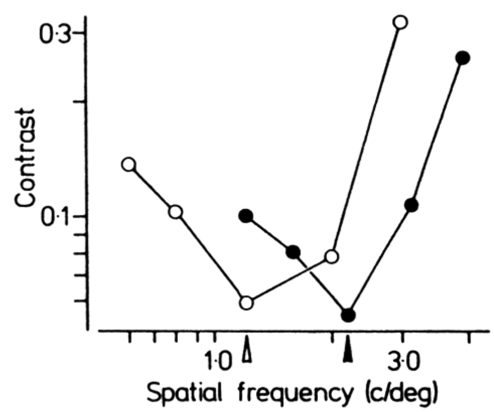

Fig. 9: The mechanisms responsible for coherent pattern perception are tuned for spatial frequency. Shown here are two tuning curves, in which contrast for coherence is measured as a function of the relative spatial frequencies of the two gratings. See text for details.

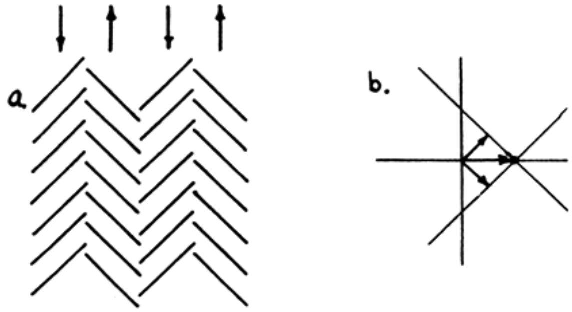

Fig. 10: The split herringbone illusion. Alternating columns of tilting line segments move up and down. Under certain conditions one has the illusion that the herringbone pattern is moving continuously to the right. See text.