Motion and Depth Perception

Although the vast majority of objects in ordinary environments are stationary the vast majority of the time, ones that move are particularly important. They include the most significant of all objects—namely, ourselves—as well as other people, animals, certain classes of human artifacts (such as cars and other vehicles), and normally stationary objects (such as rocks and trees) that are falling or have otherwise been set in motion by the action of some force. Our very survival depends critically on being able to perceive the movement of such objects accurately and to act appropriately. The consequences of failing to perceive a falling tree or an approaching car can be dire indeed!

Considering the importance of perceiving movement, it is not too surprising that the visual system is particularly sensitive to it. Indeed, objects in motion grab our attention more forcefully than almost anything else. If there is a moving object in our peripheral vision, we usually turn our eyes rapidly to find out what it is and how it is moving. This much seems obvious, but like many other obvious facts about vision, explaining how it happens is not as easy as it seems.

The first issue with which we must come to grips is the difference between the motion of an object and the motion of its retinal image. Image motion depends on both eye movements and object motion. For example, eye movements produce image motion even in a completely stationary environment; as the eye moves in one direction, the image sweeps across the visual field in the opposite direction. Here, there is image motion when there is no object motion at all. The opposite can also occur, as when a moving object is tracked by the eye: Its image on the retina will actually be stationary even though the object itself is in motion.

Thus the computational problem of motion perception is how to get from the dynamic optical event on the retina to the veridical perception of moving objects and self-motion (**egomotion**) within a generally stationary environment. This is not a trivial problem, particularly when the complications of eye, head, and bodily movements are considered. The visual system appears to solve it in at least two steps: an early process of motion analysis concerned with the registration of 2-D image motion and a later process concerned with interpreting image motion in terms of egomotion and depths of 3-D objects (either stationary or moving in 3-D space). We will consider each process in turn.

1. 2D Image Motion

1.1 The Aperture Problem

Thus far we have presented the correspondence problem as arising only in stereo and discrete motion. It seems unnecessary for real motion because when each point is displaced continuously, there would appear to be no ambiguity in the nature of the correspondence. But there is a particular version of the correspondence problem, known as the aperture problem, that arises in real motion as well. It is an important problem in motion perception not because we frequently perceive the world through apertures, but because the cells that code motion in the visual system typically have small receptive fields and therefore respond as though viewing a small portion of the visual field through an aperture.

The aperture problem refers to local ambiguity in the direction and speed of motion whenever the portion of the image that is visible within a restricted region—referred to as the aperture—lacks unique points whose correspondences can be unambiguously determined. Consider, for example, a straight horizontal line moving diagonally behind a circular aperture as shown in Figure 1A. An observer viewing such a display can know the direction and speed of motion only if the correspondence of points along the line is known. But because none of the points along the line are visually distinguishable, there are no unique points whose correspondences over time can be unambiguously determined. The center point at time t1, for instance, could correspond to the center point at time t2, but it could also correspond to any other point along the line at t2, as indicated in Figure 1A. The line's motion is therefore ambiguous. It could be specified by any vector whose tail lies along the visible line at t1 and whose head lies along the line at time t2. In fact, when people are shown a line moving behind such an aperture, they always perceive the line as moving perpendicular to its orientation (in this case, vertically, as shown in Figure 1B). This perception minimizes the speed and distance the line appears to move.

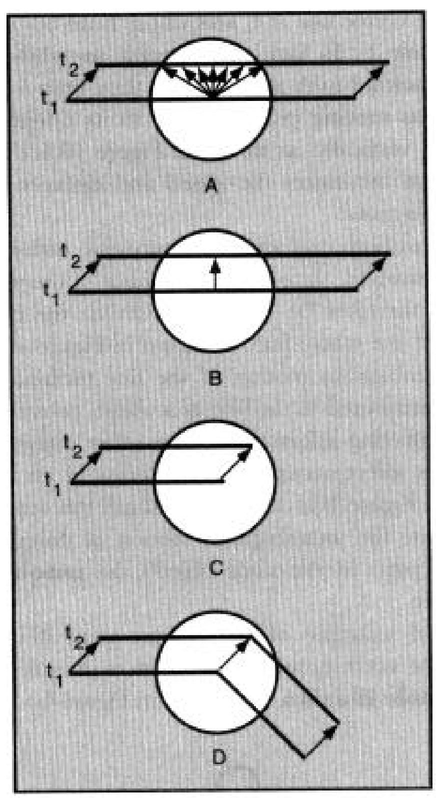
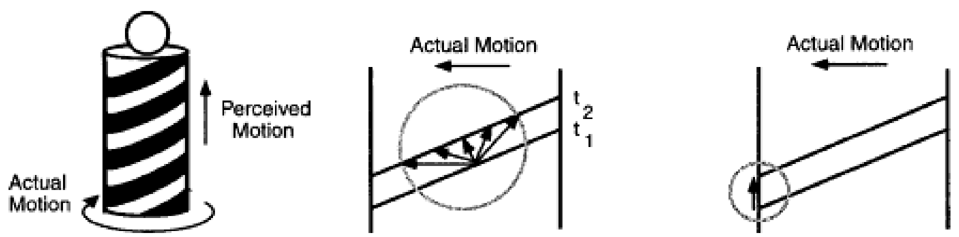


Figure 1 The aperture problem. The motion of a continuously moving straight line behind an aperture is ambiguous because any point along it at time t1 can correspond to any point at time t2 (A). In fact, observers perceive perpendicular motion (B) in the absence of any unique points. When unique points are visible (C and D), the actual motion will be perceived.

If one or both ends of a line are visible within the aperture, however, their correspondence is unambiguous and can therefore be used to determine the perceived motion of the whole line, as shown in Figure 1C. The unambiguous motion of the line terminators can then be attributed to the line as a whole, provided there is no conflicting information from other unique points. The same effect occurs with the vertex of an angle, as shown in Figure 1D. We will call this tendency to extrapolate the unambiguous motion of unique points to other parts of the same objects the unique-point heuristic.

A good example of the unique-point heuristic in solving the correspondence problem is provided by the barberpole illusion, illustrated in Figure 2A. A barberpole consists of a cylinder painted with a helix of red stripes on a white background. It rotates continuously around its cen tral vertical axis, so all points on its surface actually move laterally (leftward, in this case) at all times. Yet a rotating barberpole produces the illusion that the stripes move vertically up the pole. Why do we perceive this illusion rather than the actual sideways motion?



A. Barberpole Illusion B. Possible Motions in the Center C. Possible Motion at the Edge

Figure 2 The barberpole illusion. The stripes on a barberpole are perceived to move upward even though the pole is actually turning laterally (A). This illusion arises because motion in the center of the stripes is ambiguous (B) and motion at the sides appears to be upward (C). (See text for details.)

First, notice that along the middle of the stripes, there is directional ambiguity due to multiple solutions of the correspondence problem. Figure 2B shows the edge of just one stripe at two different times, labeled t1 and t2. From this midstripe information alone, the edges could be moving horizontally leftward (as they actually do), diagonally up and leftward, directly upward, upward and rightward, or any direction in between. But now consider the "ends" of the stripes. (They are not actually physical ends of the stripes, of course, but only the points at which the curvature of the cylinder occludes the continuation of the stripe on the other side of the pole.) These perceived ends appear to be unique points moving upward, and so their perceived upward motion specifies that the stripe as a whole is moving upward, as illustrated in Figure 2C.

This perception is illusory because the "endpoints" of the stripes are not actually the same from one time to another: They move laterally around the pole to locations that are occluded at time t2. When the correspondence arrived at by the unique-point heuristic is erroneous, the motion that is perceived will be correspondingly erroneous.

By this account, if the stripes had a clearly visible texture—for example, if they were made up of many red dots on a white background so that correspondences over time were unambiguous—the barberpole illusion should disappear. And indeed it does, provided that the texture is coarse enough.

Notice that the ends of the stripes at the bottom edge of the barberpole are actually moving leftward. Because they are true endpoints (rather than apparent ones, as are those along the sides), they specify the true motion of the pole. Why do these bottom endpoints not determine the perceived motion of the whole barberpole? The answer appears to be that they are simply overwhelmed by the larger number of ends along the sides. If barber-poles were very short and fat, the perceived motion of the stripes would indeed be horizontal. This interesting fact was discovered by psychologist Hans Wallach using a 2-D analog of the barberpole illusion in which diagonal stripes were moved behind a rectangular aperture. Regardless of the actual motion of the stripes, when they were moved behind a vertically oriented rectangular aperture, the stripes appeared to move upward. When the aperture was horizontally oriented, they appeared to move leftward.

There is a related phenomenon involving the perceived motion of "plaid" gratings created by superimposing two square wave gratings at different orientations. If a single grating is presented in motion, its perceived direction of movement is always perpendicular to the orientation of the stripes, as explained above in discussing the aperture problem. If a second moving grating of the same spatial frequency and contrast is added to the first so that it moves in a different direction (Figure 3A), the combined grating pattern is perceived to move in neither of the directions of the component gratings. Rather, it is seen to move in a direction midway between the two component motions. This result corresponds to the movement of the points where the gratings cross, as illustrated in Figure 3A. This is what would be expected from the above analysis of the aperture problem if the crosspoints are taken as unique points whose motion specifies the motion of the entire plaid pattern.

Interestingly, if the two component gratings have very different spatial frequencies (see Figure 3B) or very different contrasts, they are perceived as moving quite independently in different directions, sliding past each other . That is, they do not form a single plaid grating that moves in the direction of the crosspoints, as would be expected from a simple analysis of unique points. In fact, the crosspoints do not affect motion perception in this case at all, as though the motion system doesn't even "see" them. This may actually be the case, since different populations of cells in area V1 would respond to the motion of the two different spatial frequencies. The existence of seemingly unique points in the stimulus thus cannot affect perceived motion unless the visual system responds to them as unique points. When such points arise from information processed in two different spatial channels, as when the difference in spatial frequency of components in a plaid grating is large, the motion system does not respond to them.

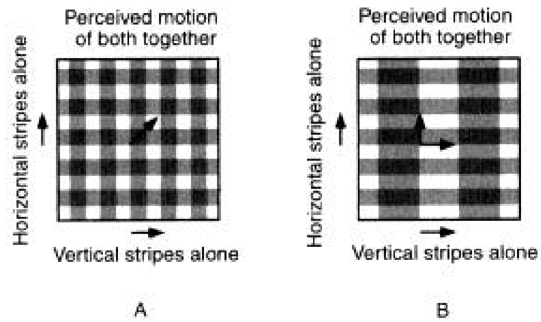


Figure 3 Moving plaid patterns. When the stripes of plaid patterns are similar in spatial frequency and contrast, they are seen to move as a single uniform pattern along the diagonal path of the intersections (A). When the gratings differ greatly in spatial frequency, they are seen to slide past each other, moving in two different directions (B).

2 3-D Motion and Depth Perception

2.1 Object Motion

The theories of motion that we have considered thus far are concerned with movement in the 2-D image: how changes in the retinal distribution of luminance can be detected over time. But this is only the first step in motion perception, for what is ultimately needed is information about how objects are moving in the 3-D environment. For a predator trying to catch its fleeing prey—and for the prey trying to escape its attacking predator—detecting motion in the 2-D retinal image is wholly inadequate. To be useful, image motion must be interpreted to provide information about the motion (or the lack of it) in environmental objects. This requires the separation of the movement of physical objects in the environment from those due to eye, head, and bodily movements. Only then does the observer have access to evolutionarily useful information about how objects and self are moving in the world.

2.2 Self-Motion and Optic Flow

Every time we change the position of our eyes or head, retinal motion is produced. This motion is global in the sense that it occurs over the entire visual field rather than differentially for selected objects. Figure 4 shows a very simple example in which the trajectories of a subset of points on visible surfaces have been tracked for a short period of time. As you can see, the visual stimulus is rich and complex.

Because most people spend a large portion of their time in locomotion, actively exploring their environment by moving their bodies, heads, and eyes, the visual system must deal with global patterns of optic flow much of the time. Optic flow is thus not a laboratory curiosity, but the dominant stimulus for normal everyday vision. The analysis of optic flow produces not only the perception of the surfaces that make up the visible environment, but also the perception of our own path through that environment. We thus use optic flow to navigate through the world and control our actions within it.

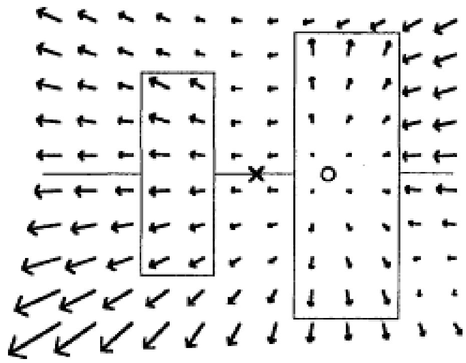
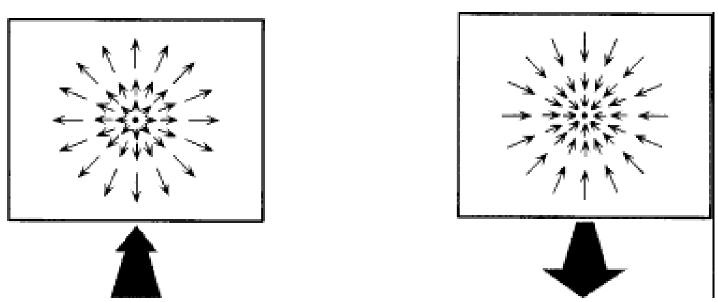


Figure 4 An example of optic flow. If an observer is moving toward the X while fixating his or her gaze on the O, the illustrated pattern of optic flow will result in an environment consisting of a ground plane, a distant wall, and two rectangular surfaces at different distances.

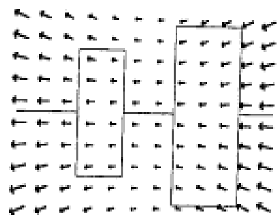
The pattern of optic flow on the retina provides information about how you are moving within your environment. We will consider a particularly simple example. If you look directly forward and walk toward a textured wall in front of you, the pattern of optic flow is that of radial expansion outward from the fixation point, as illustrated in Figure 5A. If you back away from the fixation point instead, the flow field that is created is one of contraction toward the focal point, as illustrated in Figure 5B. (These same flow fields would be generated by the swinging room moving toward or away from the observer as he or she looked straight ahead.) Flow fields get more complex when the environment contains many different surfaces, but as long as we are moving forward, they will have some expansionlike pattern as surfaces get closer.



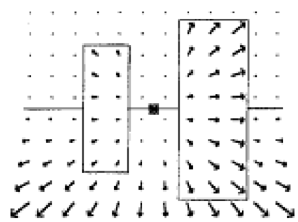
A. Motion Toward B. Motion Away

Figure 5 Optic expansion and contraction. If an observer moves closer to a surface (or the surface moves closer to the observer), its texture expands optically. Movement away results in the opposite pattern, optic contraction.

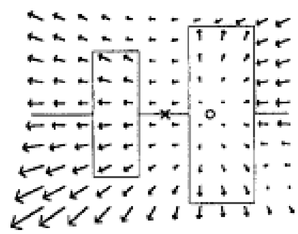
When you look in a different direction, as you also frequently do, the situation becomes more complicated, because the pattern of optical flow changes every time you make an eye movement. In the absence of other forms of image motion, an eye rotation in any particular direction causes optic flow in the opposite direction, as illustrated in Figure 6A. When a pursuit eye movement is made during self-motion—for example, to track a sign as it moves in the visual field while driving—this eye rotation component of motion is added to the flow field caused by the forward translation of the observer (Figure 6B), resulting in the complex combined optic flow depicted in Figure 6C. Here X is the direction of heading and O is the tracked point; people are able to correctly perceive their heading (toward X) as being unchanged by the eye movements that produce fixation on O. How is this possible?



A. Flow field from rightward eye rotation



B. Flow field from moving toward X.



C. Flow field from moving toward X while tracking O.

Figure 6 Combined flow field from self-motion plus eye rotation. Part A shows the flow field generated by a rightward eye rotation without self-motion, part B shows the flow field generated by self-motion toward the X without eye rotation, and part C shows the flow field generated by self-motion toward the X while the observer makes a rightward eye movement to track the O.

There are two different kinds of explanation about how people might be able to perceive their heading accurately in the face of eye movements. The retinal image theory assumes that the visual system itself is able to factor out the translational component that the eye movement produces from purely image-based information. Formal analyses have shown that optic flow on the retina contains sufficient information to do so (e.g., Longuet-Higgins & Prazdny, 1980). The extraretinal theory assumes that information about the eye movement itself—either from the eye command signal or from proprioceptive feedback—is taken into account by effectively subtracting its contribution to the flow field.

2.3 Depth and Motion

In the previous chapter, we considered how the visual system is able to use information in the 2-D image to recover the third spatial dimension of depth. Many of these factors were quite independent of motion—for example, accommodation, convergence, binocular disparity, perspective projection, texture gradients, edge interpretation, shading, and so on—but several were derived directly from motion.

Rigid Motion in Depth

In studying the perception of motion in simple 2-D displays consisting of pairs of continuously moving points of light, Swedish psychologist Gunnar Johansson discovered some powerful depth effects in motion perception. In one display, two points moved back and forth in synchrony while they also moved closer together and farther apart, as shown in Figure 7A. You might expect that this simple pattern of retinal motion is what observers would perceive, but most did not. Instead, they saw the two dots moving forward and backward in depth, as though they were lights attached to the ends of a rigid vertical rod (Figure 7B).

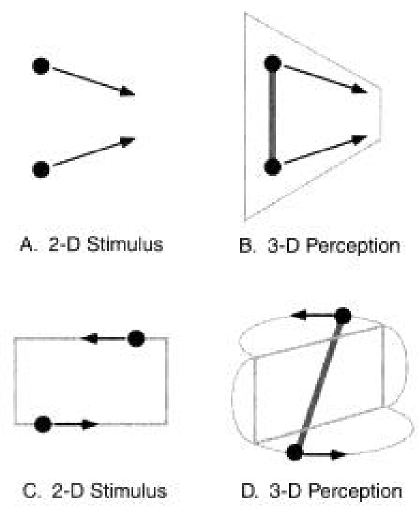


Figure 7 Depth effects in perceived motion of simple 2-D events. Two dots moving in phase as shown in part A are perceived as two dots moving in depth as though attached to the ends of a rigid rod (dashed line), as illustrated in part B. Two dots following each other around a rectangular path as shown in part C are also seen to move in depth as though attached to the ends of a rigid rod pivoting through the center, as illustrated in part D.

Other displays produced similar depth effects. Whenever the pattern of 2-D motion could be perceived as rigid motion in depth, it was. This included several rather surprising cases, such as the one shown in Figure 7C. The display consisted simply of two lights traversing the same rectangular path on opposite sides, but this is not what people perceived. Rather, they saw the lights moving as though attached to the ends of a rigid rod that rotated in 3-D space, producing the complex motion pattern in depth depicted in Figure 7D.

Such findings suggest that there is something quite special about rigidity in perceiving object motion: All else being equal, if there is an interpretation in which rigid motion can be perceived, it will be. We will call this the rigidity heuristic. We will see that it plays an important role in understanding the relation between depth and motion, although there are cases in which other factors override it.

Speed of Self-Motion

Given that the direction of self-motion can be determined visually, the other component that is required to specify one's motion through the environment is speed. Information about absolute speed is not available solely from optic flow because speed is defined as the rate of change in location over time, and this requires information about absolute distance. Absolute distance cannot be determined from purely optical information without additional assumptions because of the fundamental scaling indeterminacy: The same retinal events could result from environmental events in which all environmental distances were multiplied by a constant factor. Determining absolute speed from optic flow thus requires some scaling information from a nonoptical source, such as eye convergence or familiar size.