SECTION
THREE

MECHANISMS FOR MOTION PERCEPTION

BY EDWARD ADELSON

One of the goals of vision science is to understand the workings of visual systems from the diverse perspectives of psychophysics, neurophysiology, and computation, and to develop this understanding into an integrated theory of vision. This program has been particularly successful in the study of motion perception.

Let us begin by considering one of the long-standing problems of motion perception: the phenomenon of apparent motion, by which a rapid sequence of movie frames gives rise to the impression of motion. For many years, this was a hotly debated phenomenon. But recently a unified view has emerged that combines insights from several converging lines of investigation. In this view, motion is conceptualized as a kind of slant or orientation in space-time. Any system that is designed to detect this space-time orientation will inevitably experience the "illusion" of apparent motion in the right conditions From this perspective, it is no mystery that we see motion in movies, and it is probably misleading to think of the effect as an illusion.

To illustrate the spatio-temporal approach, consider a scene containing a moving object, such as that shown in Figure 1 (adapted from Ref. 1). In Figure 1a, a vase is seen to move in a rightward direction. A movie of this scene would consist of a sequence of vase images in which the vase is successively displaced to the right. We could stack all of these frames up into a solid, as illustrated in Figure 1b, making something like a flip-book; a skeleton view is shown in Figure 1c. We are using the third dimension to represent time, and since images contain only the two spatial dimensions (x,y), the axes of our spatio-temporal representations are (x,y,t).

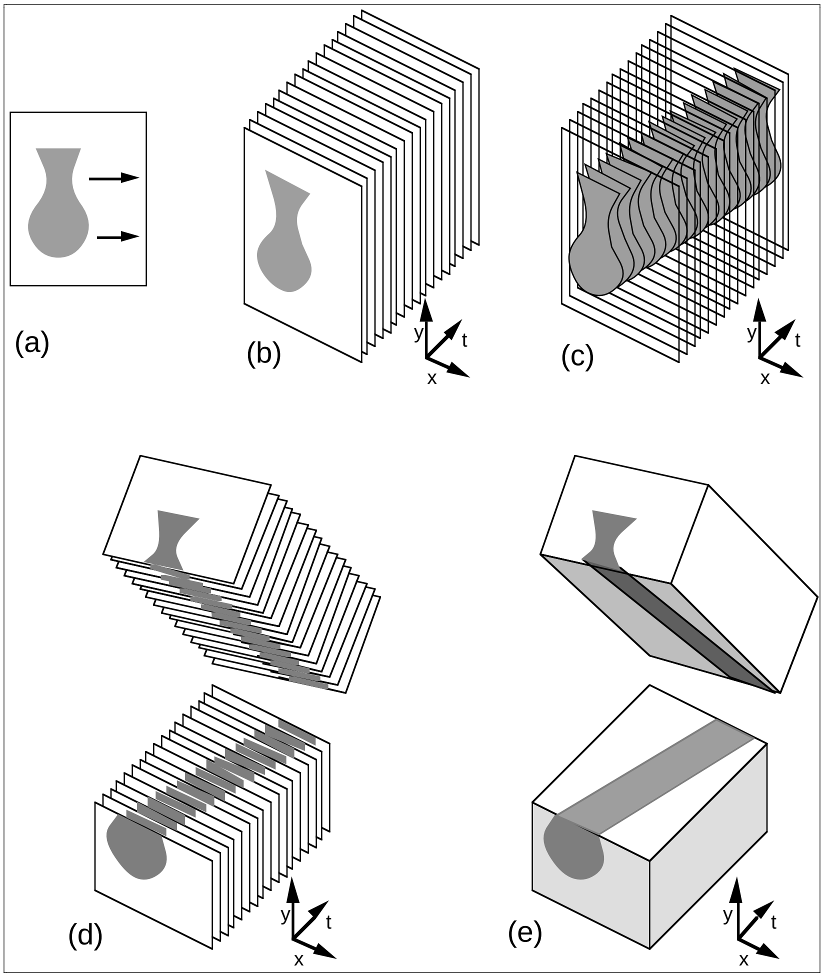

FIGURE 1. a: An image of a vase moving to the right. b: A sequence of frames may be piled up to form a flip book; time is the third dimension. c: A skeleton view of the (x,y,t) volume helps suggest its structure. d: The space-time volume may be sliced to illustrate the fact that the motion is equivalent to spatio-temporal orientation. e: In the case of continuous motion, the volume is densely filled. The moving vase traces out an extruded shape that is sheared due to the motion.

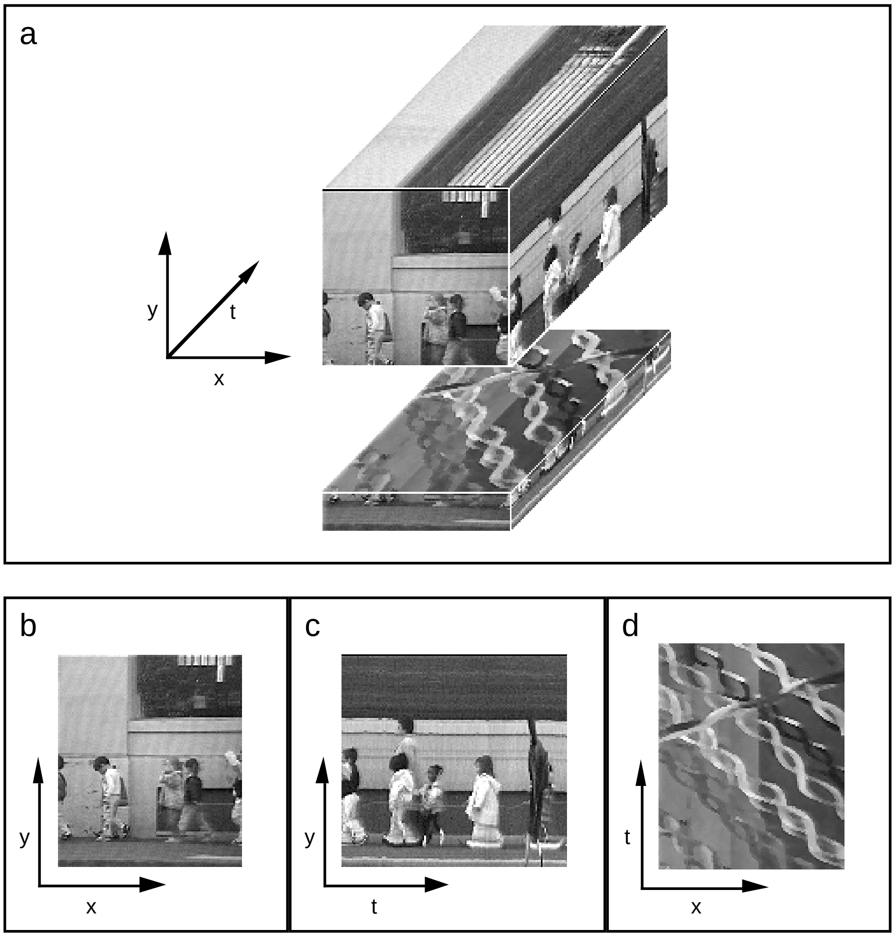

FIGURE 2. a: An (x,y,t) volume derived from a video sequence. The volume has been sliced along an (x,t) plane to reveal the internal structure. b: An (x,y) slice, corresponding to the front face of the volume. It shows an ordinary video frame of children walking. c: A (y,t) slice. The children are visible as they scan out their own images by working past, a fixed vertical line d: An (x,t) slice taken at ankle height. The ankles trace out character istic "braided" patterns.

To get a better understanding of the structure of the space-time volume, we can cut a slice through it in an (x,t) plane, as illustrated in Figure 1d. The vase traces out an extruded shape that is sheared due to the motion. In the case of continuous motion, the spatio-temporal volume is densely filled, as shown in Figure 1e. The (x,t) slice is slanted as a result of the rightward motion.

Figure 2a shows an (x,y,t) volume taken from a video sequence showing a group of children walking in front of a building. The volume has been split along the (x,t) plane at ankle height so that we can see the spatio-temporal patterns traced out by the children as they walk. Figure 2b shows a single frame from the video sequence, which is to say, a single (x,y) slice. Figure 2c shows a (y,t) slice. It corresponds to the information one would see through a fixed vertical slit over an interval of time. As the children walk past the slit, they draw their own portraits, with some distortion since different parts of their bodies are moving at different speeds. Toward the end of the sequence, an adult walks past the children in the opposite direction; since she is moving more quickly, her portrait appears skinnier.

Figure 2d shows an (x,t) slice at ankle height. The children produce marvelous braided patterns as they move their feet. Note that the adult's path can also be seen toward the top of the slice. Since the adult walks faster, the path is slanted closer to horizontal.

The problem of analyzing velocity is analogous to the problem of analyzing orientation, except that the orientation is embedded in a volume of space-time rather than ordinary space. From a mathematical point of view, the problems of motion analysis and orientation analysis are very much the same. Since the basic problem of spatial orientation analysis has been well-studied, it is useful to consider what is known in the spatial domain and then generalize it to include the space-time.

When Hubel and Wielsel embarked on their Nobel Prize winning research on neurons in visual cortex, they found that most of the cells responded best to the elongated structures such as edges or bars, and that each cell had its own preferred orientation. The orientation preference corresponded to the shape of the "receptive field," which is the pattern of positive and negative excitation that the cell displayed when stimulated in different portions of the visual field. Figure 3a shows an example of a receptive field from a "simple cell," containing one excitatory region and one inhibitory region, side by side. This type of neuron responds well to vertical edge-like patterns. A similar oriented receptive field, which responds well to bar-like patterns, is shown in Figure 3b.

Consider the responses of a neuron with a receptive field like that of Figure 3a. With the vertical light-dark edge of

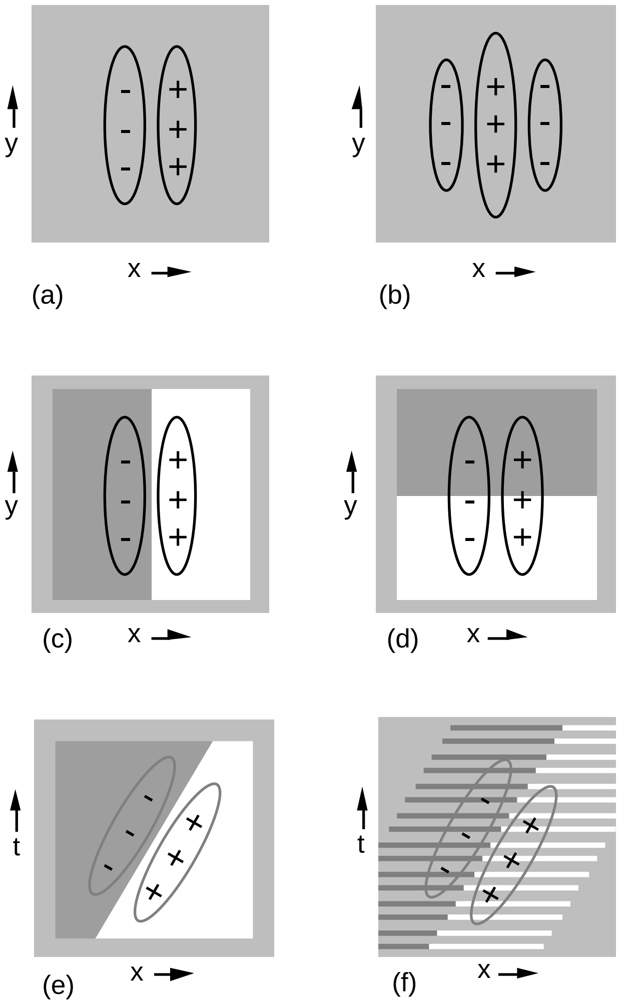

FIGURE 3. a: The receptive field of a cortical neuron that responds well to vertical edge-like structures. b: The receptive field of a neuron that responds well to bar-like structures. c: A vertical edge aligns with the excitatory and inhibitory regions of the receptive field, leading to a forge response. d: A horizontal edge produces equal responses in the excitatory and inhibitory regions, leading to little or no net response. e: An oriented receptive field prefers edges of the some orientation. In this case, the axes are (x,t), so the orientation refers to motion. f: A neuron that responds well to continuous motion will also respond well to sampled motion, such as a seen in a movie.

Figure 3c, the excitatory and inhibitory regions line up with the light and dark regions of the stimulus, leading to a strong response. On the other hand, the neuron will respond poorly to an edge at a perpendicular orientation—such as the one shown in Figure 3d—since the responses in the excitatory and the inhibitory regions average to zero.

Figure 3e shows a receptive field that is tuned for slanted edges. But note an additional change here: the axes are labeled (x,t) rather than (x,y). That is, this hypothetical neuron has a spatio-temporal receptive field that will respond preferentially to edges moving in a certain direction with a certain speed. And not only will it respond well to continuous motion; it will also respond well to the sampled motion of a movie sequence, as is shown in Figure 3f.

The notion of a spatio-temporal receptive field is more straightforward than it may seem at first. In the case of a cell showing linear summation, the receptive field is simply a weighting function that describes how pointwise inputs at various positions and times are summed to generate the current output. It turns out to be simple to generate receptive fields with spatio-temporal orientation using physiologically plausible mechanisms.[1,2]

This manner of thinking about motion sensitivity originated in studies of human psychophysics,[1-3] but it led physiologists to look for such spatio-temporal receptive fields in directional sensitive neurons. The search has met with success: several investigators have reported finding cortical neurons that behave in a manner similar to that predicted by the spatio-temporal models.[4-6] The spatio-temporal approach has also proven to the useful in the domain of machine vision.[7-8]

One style of spatio-temporal model, proposed by Adelson and Bergen,[1] combines the outputs of receptive fields of odd and even symmetry to produce a phase-independent measure of motion energy. Such a mechanism (similar to a mechanism proposed for spatial vision[9]) gives a positive response to a moving bar or edge, regardless of the contrast sign of the stimulus, as long as it is moving in the preferred direction. The predicted behavior is similar to that observed in complex cells in visual cortex.[10]

One can also design a model in which two such neurons (e.g., a leftward and a rightward neuron) are in push-pull opposition. These opponent-energy models can be equivalent to a form of motion mechanism known as a Reichardt model,[3,11] which was originally proposed for the visual system of the fly. However, cells in mammalian cortex do not appear to behave as predicted by opponent energy or Reichardt models in their classic form.[10]

To analyze the three-dimensional (x,y,t) structure of image sequences, we must use three-dimensional filters. Figure 4 (and the cover) shows, in a schematic form, a set of filters that can be used in spatio-temporal analysis. The alternating ellipsoidal shapes represent positive and negative values above or below an arbitrary threshold. These particular filters are idealized ones called steerable filters. It is possible to synthesize a steerable filter of any intermediate orientation by taking a linear combination of a small number of

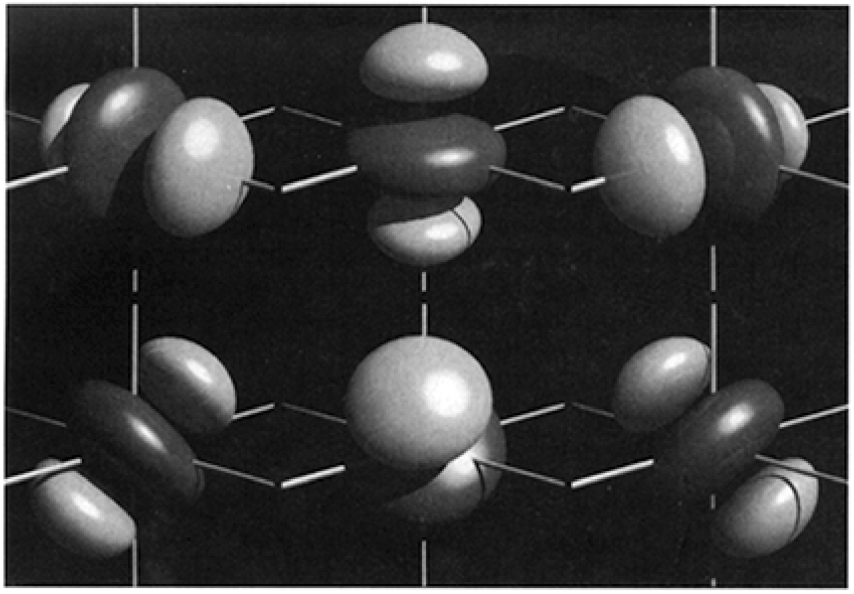

FIGURE 4. A set of "steerable" filters in three dimensions. If the dimensions are (x,y,t) then the filters can be used to analyze motion.

basis filters; this turns out to be useful in computational systems for analyzing orientation and motion.[12]

OTHER DIMENSIONS

As we have seen, the measurement of motion is related to the measurement of spatial orientation by a simple change of variables. The same notion can be extended to other dimensions as well. Figure 5 shows the same slanted stimulus with four different choices of dimensions. Figure 5a shows the ordinary case of space and Figure 5b shows the case of space-time just discussed, where orientation becomes motion. Figure 5c shows a case of (t, λ), that is, time and wavelength. This temporo-chromatic image corresponds to a change in color and intensity over time. Figure 5d shows an example where one axis is viewing position in the x direction (V_x), and the other axis is spatial location in the image (x). This leads to a plot of parallax: as the viewing position is moved, the position of the feature in the image also moves at a certain rate, where the rate depends on the object's distance from the viewer. Thus the slant gives an estimate of depth. Humans can gather this information either by moving the head (motion parallax) or by using the two samples of spatial position acquired by the two eyes (binocular disparity). The two binocular samples are indicated by the two vertical lines.

One can formalize this approach by defining a seven-dimensional function called the "plenoptic function".[13] The basic measurements of early vision—including orientation, color, motion, flicker, disparity, etc.—can be considered as measurements of local change within this function. The same computational issues that arise with motion are found with these other measurements as well. Moreover, one can think of many neurons in visual cortex as possessing receptive fields tuned to change along various axes in plenoptic space. This point of view leads to some interesting connections. For example, color-opponent cells are similar to flicker-sensitive cells under a change of variables.

HIGHER LEVEL MECHANISMS

The simple mechanisms of early vision are only the beginning of the story. There are many other, higher-level measurements that involve more stages of processing. In the case of motion analysis, we know that there are important mechanisms beyond those sensitive to motion energy. The first clear evidence for multiple types of motion mechanisms was provided by Braddick[14] in his now classic experiments with random dot stimuli. He distinguished between "short range" and "long range" motion phenomena. The short-range phenomena can probably be understood in terms of the motion energy mechanisms we have just discussed. The long-range phenomena require

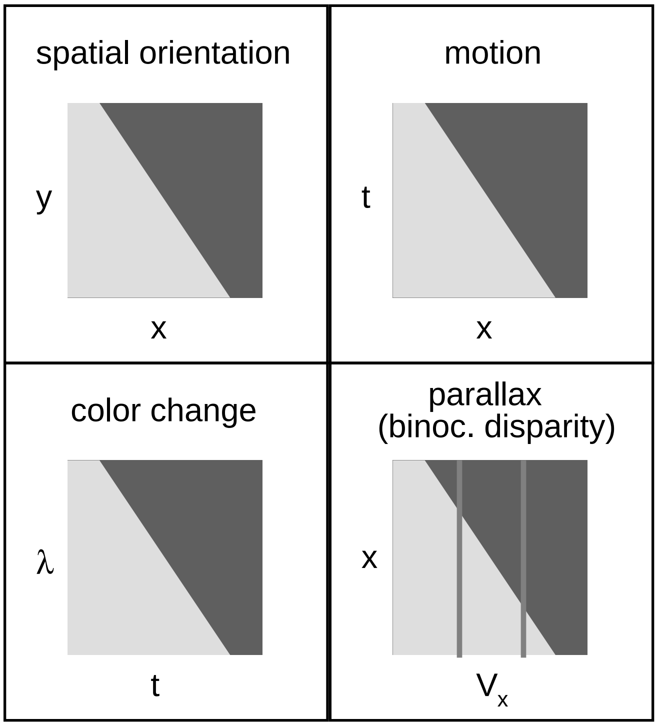

FIGURE 5. a: A tilted edge in (x,y). b: The same tilted edge in (x,t) corresponds to motion. c: A tilted edge in (t,λ) corresponds to color change. d: A tilted edge in the (V_x,x) domain corresponds to parallax, indicating the object's distance from the viewer.

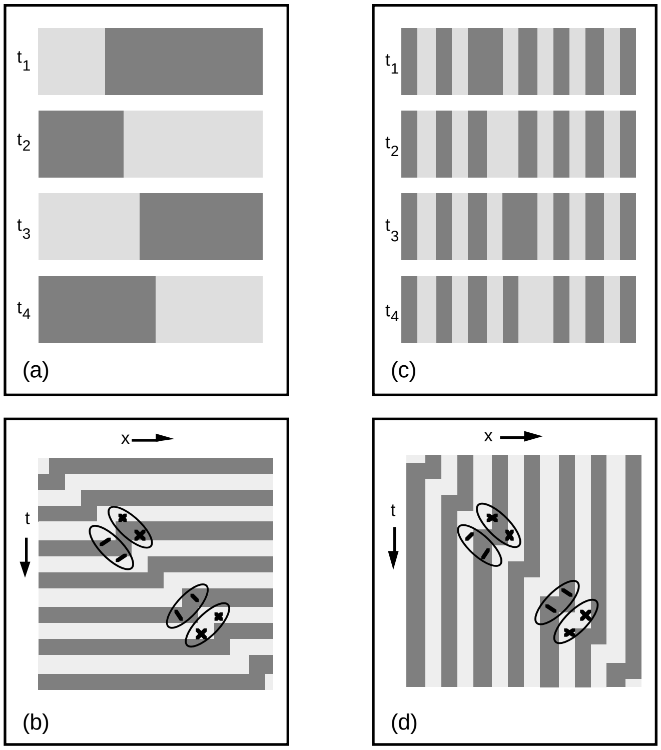

FIGURE 6. a: A sequence of four frames showing an edge moving to the right while alternating contrast sign. b: An (x,t) plot of the same sequence. c: A sequence of four frames in which a fat bar moves to the right and alternates sign. d: An (x,t) plot of the same sequence.

other explanations.

There is a wide range of evidence for the existence of at least two types of motion mechanism (see, for example, the review by Anstis[15]). One of the characteristics of the shortrange or motion-energy mechanisms is that they produce a strong "motion aftereffect." If one stares at a moving pattern for a minute or so and then looks at a stationary pattern, the stationary pattern seems to move in the opposite direction. For example, after one gazes at a waterfall, and then looks to the rocks beside it, the rocks appear to drift upward.

It is possible to make a stimulus that appears to move rightward, although it contains no rightward motion energy. A simple example used by Anstis and Cavanagh[16] is a flickering edge that is moving in a rightward direction, as illustrated in Figure 6a. In each frame in the sequence, the edge moves to the right and reverses contrast. The frames labeled t_1, t_2, t_3, and t_4 are examples of four such images. The stimulus appears as one might expect: it is an edge moving to the right while undergoing a flicker at the same time.

To understand how basic motion detectors will respond to this stimulus, we can make an (x,t) plot, as shown in Figure 6b. The plot consists of a set of horizontal bars with a "kink." The orientation of the kinked region corresponds to the rightward motion. However, the receptive field that is lined up with the rightward motion (shown toward the top), actually gives a zero response, since it receives the same stimulation in both its excitatory and its inhibitory regions. On the other hand, the receptive field that is lined up with the reverse direction (shown toward the bottom) gives a positive response, because its excitatory and inhibitory receptive fields line up with the light and dark regions of the stimulus.

The stimulus is paradoxical: it moves rightward and is seen to move rightward, and yet the neurons tuned for the motion energy should see leftward motion. There is one other interesting fact: The stimulus produces a motion aftereffect that is in the same direction as the rightward motion, rather than in the reverse direction as it would normally be.

The explanation seems to be this: the simple motion detectors that respond to motion energy are unable to see the rightward motion in the flicker stimulus, but there is some higher level mechanism—involving a more complex combination of form and motion processing—that is able to correctly sense this motion. At the same time, the basic detectors are busily responding to the stimulus, even if their response is ultimately overridden by the higher-level processes. The basic detectors are responsible for the motion aftereffect, which goes in the direction opposite the motion that they silently detected.

In another experiment, Mather *et al.*[17] presented a series of alternating white and dark bars, as shown in Figure 6c. On each frame, one of the bars was fat. The fat bar moved to the right on successive frames and also alternated sign between light to dark. This stimulus, like the flickering edge, appeared to move to the right, but generated a motion aftereffect in the rightward direction, which again is opposite of the normal aftereffect. The (x,t) analysis, shown in Figure 6d, reveals that this stimulus is actually the same as the flickering-edge stimulus, except that the x and t axes are exchanged. As before, there is no motion energy in the direction of perceived motion, but there is strong motion energy in the reverse direction. Apparently it is the adaptation of the silently responding motion detectors that is causing the motion aftereffect.

Chubb and Sperling[18] have developed some clever mathematical methods to design spatio-temporal stimuli that are completely balanced in terms of their leftward and rightward motion energies, and which therefore can be expected to produce equal responses from rightward and leftward units tuned for motion energy. When a subject perceives motion with these stimuli, the percept must originate in a source other than the basic motion detectors. A great variety of these stimuli can produce a strong impression of motion. But these stimuli produce little or no motion aftereffect.

An interesting experiment by Hochberg[19] indicates that mechanisms of increasing sophistication come into play over longer time scales. Hochberg presented a row of geometrical shapes, such as circles and squares, as illustrated in Figure 7. On each frame, the entire row jumped to the right by some distance. The perception of motion in such a stimulus normally follows a "nearest neighbor" rule, which is the same rule predicted by a motion-energy mechanism. If the row jumped only slightly—as shown in the transition from Figure 7a to Figure 7b—then the motion was seen correctly. If the row moved too far, however, the nearest neighbor rule (and the motion energy calculation) would predict a backward motion. For example, if the circle Figure 7c jumped so far that it was almost lined up with the old position of the square—as in Figure 7c—then the percept was of a leftward motion rather than a rightward motion. There is nothing unexpected in this observation.

But here is the surprise: when Hochberg slowed down the presentation rate, the perceived motion was strongly influenced by the shapes. In the case of Figures 7c and 7d, the motion was seen going to the right, even though the nearest neighbor rule predicts motion to the left. At these slower rates, the motion path was the one that maintained object identity. Thus it seems that there do exist motion mechanisms sensitive to the more sophisticated properties of the stimuli, but they take a bit of time to come into full play.

The slower long-range mechanisms are probably not important in the perception of motion in rapid sequences such as movies or in the continuous motions seen in ordinary life. These mechanisms may have evolved to help observers piece together a scene from a series of glimpses, such as when we move our eyes in saccades from one point to another.

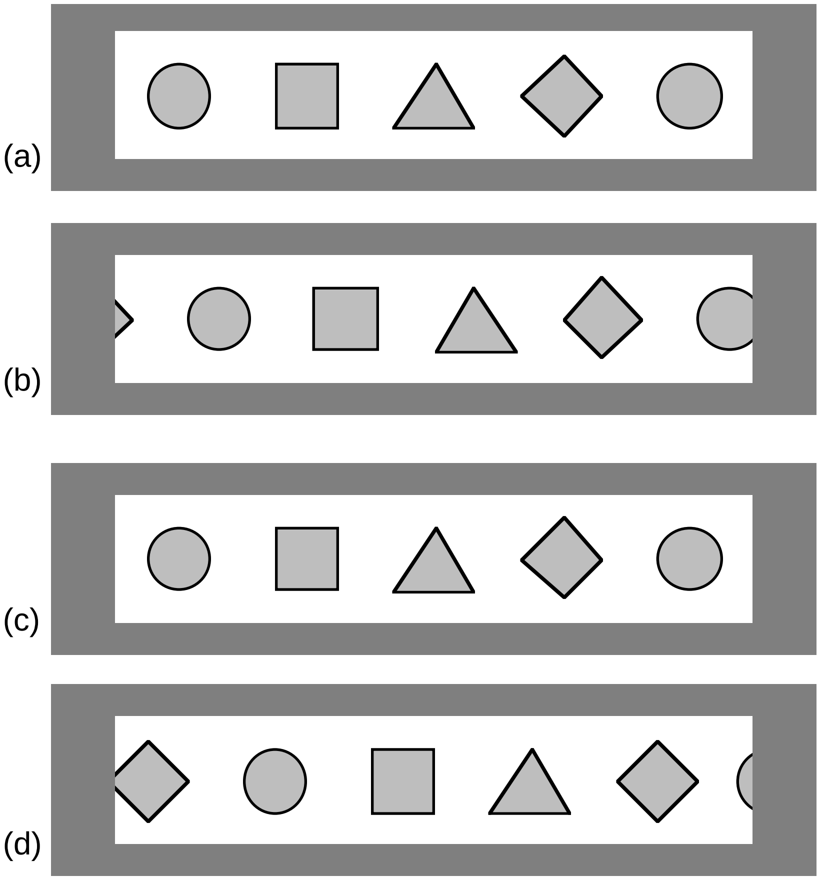

FIGURE 7. a,b: Two frames of a sequence shown by Hochberg. Motion follows a nearest neighbor rule. c,d: Two frames in which the nearest neighbor rule operates at high presentation rates, but a shape-based matching operates at lower presentation rates.

CONCLUSIONS

Motion perception can occur with a variety of stimuli, and there may be several different mechanisms involved in motion analysis. The simplest sort of motion stimulation can be considered to involve patterns that are oriented in space-time, and which possess local "motion energy." The simplest motion mechanisms are those that respond to the motion energy in the stimulus. Many phenomena in human motion perception can be explained in terms of these mechanisms, including the apparent motion seen in a rapid sequence such as a movie.

Moreover, recent physiological findings indicated that there are neurons selectively tuned for motion energy in the striate cortex. But these cells are not only the ones involved in motion perception, since humans can perceive motion involving spatio-temporal structures that lack motion energy in the direction of perceived motion. To detect this higher-order motion, there must be mechanisms involving more complex processing; some of these systems also appear to operate over longer times and distances than the basic motion energy mechanisms.

ACKNOWLEDGEMENTS

This work was supported in part by a contract with SECOM Intelligent Systems Laboratory, Japan, and by NSF Grant IRI 871-939-4.

REFERENCES

1. E.H. Adelson and J.R. Bergen, "Spatiotemporal energy models for the perception of motion," J. Opt. Soc. Amer. **A 2**, 1985, 284-299.
2. A.B. Watson and A.J. Ahumada, "Model of human visual-motion sensing," J. Opt. Soc. Amer., **A 2**, 1985, 322-342.
3. J. van Santen and G. Sperling, "Elaborated Reichardt detectors," J. Opt. Soc. Amer. **A 2**, 1985, 300-321.
4. R.C. Reid *et al*., "Linear mechanisms of directional selectivity in simple cells of cat striate cortex," Proc. Natl. Acad. Sci. USA, **84**, 1987, 8740-8744.
5. J. MacLean and L.A. Palmer, "Contribution of linear spatiotemporal receptive field structure to velocity selectively of simple cells in area 17 of cat," Vision Res., **29**, 1989, 675-679.
6. D.B Hamilton *et al*., "Visual cortical receptive fields in monkey and cat: spatial and temporal phase transfer function," Vis. Res. **29**, 1989, 1285-1308.
7. D.J. Heeger, "A model for the extraction of optic flow," J. Opt. Soc. Amer. **A 4**, 1987, 1455-1471.
8. R.C. Bolles *et al*., "Epipolar-plane image analysis: an approach to determining structure from motion," Int. J. Computer Vision, **1**, 1987, 7-55.
9. G. Granlund, "In search of a general picture processing operator," Computer Graphics and Image Processing, **8**, 1978, 155-173.
10. R.C. Emerson *et al*., "Movement models and directionally selective neurons in the cat's visual cortex," Soc. Neurosci. Abstr. **13**, 1987, 1623; R.C. Emerson *et al*., "Directionally selective complex cells and the computation of motion energy in cat visual cortex," (manuscript under review).
11. W. Reichardt, "Autokorrelationsauswertung als functionsprinzip des zentralnervensystems." Zeitschrih Naturforschung, **12b**, 1957, 447-457.
12. W.T. Freeman and E.H. Adelson, "Steerable filters for early vision, image analysis, and wavelet decomposition," Proceedings Int. Conf. on Computer Vision, Osaka, Japan, 1990; W. Freeman and E.H. Adelson, "Steerable filters for image analysis," IEEE Trans. Patt. Anal. and Mach. Intell., (in press, 199l).
13. E.H. Adelson and J.R. Bergen, "The plenoptic function and the elements of early vision," M. Landy and J.A. Movshon (eds.), *Computational Models of Visual Processing*, MIT Press, Cambridge, Mass., 1991.
14. O. Braddick, "A short-range process in apparent motion," Vision Res. **14**, 1974, 519-527.
15. S.M. Anstis, "The perception of apparent movement," Phil. Trans. R. Soc. London Ser. B, **290**, 1980, 153-168.
16. S. Anstis and P. Cavanagh, "What goes up need not come down: moving flicker edges give positive aftereffects," J. Long and A.J. Baddeley (eds.), *Attention and Performance*, Vol. **IX**, Cambridge University Press, 1981.
17. C. Mather *et al*., "A moving display which opposes short range and longrange signals," Perception, **14**, 1985, 163-166.
18. C. Chubb and G. Sperling, "Drift-balanced random stimuli: A general basis for studying non-Fourier motion perception," J. Opt. Soc. A, **5**, 1988, 1986-2006.
19. J. Hochbergand V. Brooks, "The perception of motion pictures," *Handbook of Perception*, E.C. Carterette and M. Friedman (eds.), Academic Press, New York, N.Y., Vol. **10**, 1978.

EDWARD H. ADELSON *is associate professor of vision science with the Media Laboratory and Department of Brain and Cognitive Science, Massachusetts Institute of Technology, Cambridge, Mass.*