Motion Perception, Elementary Mechanisms

David C. Burr

Visual motion is essential for many diverse aspects of biological function, varying form rapid detection of predators and prey, navigation through the visual environment and constructing a three-dimensional visual representation from two-dimensional retinal input. However, motion information is not provided by the instantaneous retinal signal, but has to be computed from the temporal variations in luminance over the image. While the neural mechanisms that achieve this vary considerably throughout the animal kingdom, the underlining principles of the algorithms seem to be very similar.

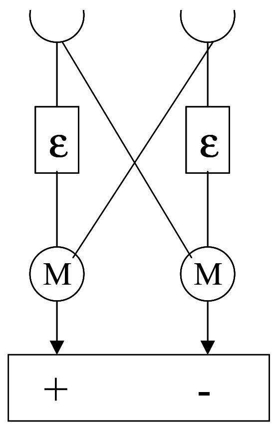

Figure 1
Simplified "full Reichardt detector" (adapted from Reichardt, 1961).

Models of motion perception

In biological visual systems motion is initially analyzed in parallel by arrays of local motion-detectors that exhibit certain basic properties: they require at least two spatially separate sampling units, one delayed with respect to the other, that are combined (usually non-linearly) to create directional selectivity. Werner Reichardt (1961) was the first to provide a formal model of a motion detector based on these principles, in what has become know as a "correlator-type" model, or more simply, the *Reichardt detector*. The detector, at its simplest, is illustrated in figure 1. The response of two spatially separated units ($\Delta\varphi$ apart) are multiplied together (at M), after one has been delayed by ε. The figure illustrates two such units arranged mirror symmetrically, using the same input. The unit on the left will respond best to rightward motion, maximally for speeds of $\Delta\varphi/\varepsilon$; that on the right responds best to leftward velocities of $\Delta\varphi/\varepsilon$. Each unit M can be considered to be an elementary motion detector, in that it shows a direction preference. However, by combining the output of two such mirror-symmetrical units (subtractively in this case) the direction selectivity is further enhanced, to produce what is referred to as the *full Reichardt detector*.

The essential components of the Reichardt detector – spatial and temporal asymmetries and cross-correlation – can be implemented in many different ways. The initial model was inspired by the fly visual system, where the two sampling points are adjacent ommatidia, and the temporal delay ε introduced by some form of delay line, typically a low-pass filter. Models of human motion have been heavily influenced by the application of Fourier analysis to vision research, showing spatial and temporal filtering of the visual input at early stages. For moving stimuli, detectors are tuned in both space and time, leading to spatio-temporally oriented filters, or receptive fields. This concept has proven invaluable, not only in constructing physiologically plausible models of motion perception, but also in explaining how the form of moving objects is encoded (Burr & Ross, 1986).

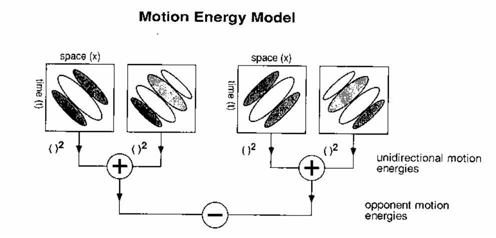

Figure 2
An example of a motion detector based on filters oriented in space-time, reproduced with permission from Adelson and Bergen (1985).

One specific example of a model based on this concept is shown in figure 2. The model starts with spatio-temporally oriented receptive fields tuned to a finite band of spatial and temporal frequencies, and hence to motion in a given direction (corresponding to a preferred orientation in the spatio-temporal plane). The orientation in space-time is readily achieved by linear combination of filters with appropriate spatial and temporal phase-shifts. In the particular model shown in figure 2, the output of two such filters in quasi-quadrature phase in space and time, is squared then summed, to produce what has been termed "uni-directional motion energy". This model responds to a drifting sinusoidal grating with a constant response, strongest when the velocity of the sinusoid corresponds to the orientation of the spatio-temporal receptive field, and weakest when in the orthogonal orientation (opposite direction). However, like the simple Reichardt detector, such a motion unit is not in itself a true motion detector, in that it will respond to many stationary transient stimuli, such as to a briefly flashed pattern of appropriate spatial frequency. Further specificity is achieved by inhibition between opponent motion energies, either by subtraction, as shown here, or by division. Interestingly, the full version of the motion energy model is formally equivalent to the full Reichardt motion detector, elaborated to include a spatial and temporal filtering stage, even though no part of the Reichardt detector corresponds to the uni-directional motion energy extractors (Adelson & Bergen, 1985).

Physiological measurements of neurons in macaque monkey visual cortex have identified plausible neural substrates for the two stages of the motion energy model (Qian & Andersen, 1994). Cells in the primary visual cortex V1 show directional selectivity, but also respond well to bi-directional motion; this is consistent with the expected performance of the first stage. However, cells in the middle temporal area (MT) show a strong inhibition by motion in the non-preferred direction, consistent with opponent motion stage of the model. FMRI studies in humans provide support for this suggestion: V1 responds more strongly to counter-phased sinusoidal gratings (that can be considered as the sum of two opposing drifting gratings) than to a single component drifting grating; whereas in MT complex, the result is reversed, with a much stronger response to the single component (Heeger et al., 1999).

Velocity tuning

The selectivity to speed of the two motion detectors of figures 1 and 2 can be varied by changing either the temporal or the spatial characteristics. For the Reichardt detector, the preferred speed can be increased either by increasing the spacing $\Delta\varphi$ between the two sampling points, or by decreasing the delay ε. Similarly, for the energy model, where the spatial and temporal offsets are given by phase shifts, preferred speed will depend on both spatial and temporal frequency preference. In

humans it is possible to measure spatial and temporal selectivity, using a variety of techniques, including "masking", where one measures contrast sensitivity to a "test" stimulus in the presence of a high-contrast "mask". The assumption is that the mask will cause maximum desensitization when its spatio-temporal characteristics match that of the detector responding to the test. To study motion perception, the test stimuli were drifting sinusoidal gratings of variable spatial and temporal frequency, displayed together with mask gratings, also varying in spatial and temporal frequency (Anderson & Burr, 1985). Over a wide range of spatial frequencies (0.025 c/deg to 15 c/deg), maximal masking occurs when the frequency of the mask matches that of the test. This suggests that there exist a battery of detectors with preferred spatial

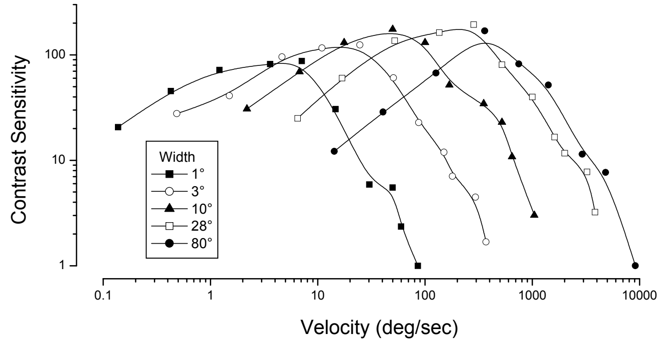

Figure 3

Contrast sensitivity for detecting the direction of motion biphasic bars of various sizes, as a function of speed (reproduced from Burr and Ross, 1982).

The lowest speed at which direction can accurately be discriminated is about 1 min/sec, for small stimuli moving over the fovea. This threshold increases steadily with eccentricity, reaching 8-10 min/sec at 90° eccentricity (largely explained by the optical degradation in the periphery). However, the upper limit of motion detection is not a fixed speed but, as may be expected from the previous paragraph, varies considerably with the spatial frequency content of the stimuli (Burr & Ross, 1982). This is brought out clearly in figure 3, showing contrast sensitivity (inverse of contrast thresholds) for biphasic bars (signal cycles of sinusoid) of various sizes, as a function of drift speed (abscissa). The small bars were seen best (required least contrast to discriminate their direction) when moving slowly, and could not be resolved at all at speeds above 100 deg/sec. The largest bars, however, were best seen when moving at 500 deg/sec, and could still be reliably resolved at 10,000 deg/sec. Thus the upper limit of motion perception is not so much a speed limit as a temporal frequency limit. The large variation in receptive field size ensures that human motion perception can operate over an extremely wide range of speeds, spanning nearly six orders of magnitude (0.015 to 10,000 deg/sec).

Apparent motion

Much of the motion we view daily at the cinema and on television is not real motion but an illusion created by displaying a series of still pictures in rapid succession (24 Hz for cinema, 60 Hz for NSTC television). This type of motion is referred to as "apparent motion", "stroboscopic motion" or, most accurately, "sampled motion". For some time it was thought that apparent motion may be detected by different processes from those detecting real motion, but recent studies find little justification for this view. Most motion detectors that incorporate spatio-temporal filtering will respond well to sampled motion, provided the sampling rate is sufficiently high. The spatio-temporal trajectory for apparent motion is a row of dots in space-time. If the spatio-temporal receptive fields (figure 2) are orientated parallel to this trajectory, they will integrate the discrete samples, effectively causing the motion to become continuous (Burr & Ross, 1986).

The minimum theoretical sampling rate is given by the Nyquist limit, that requires that the image be sampled at at least twice the temporal frequency of image motion. Sampling below this frequency will cause *aliasing*, well-illustrated by the so-called "wagon-wheel" effect: periodic moving stimuli such as wagon-wheels in Westerns are seen to stop and reverse direction as the wagon accelerates. When the repetition frequency of spokes exceeds half the sampling frequency (12 Hz for cinema), it will be undersampled, creating strong aliasing in the form of erroneous motion. The conditions under which sampled motion is indistinguishable from smooth motion can be predicted quantitatively from measurements of contrast sensitivity and linear systems analysis(Burr, Ross & Morrone, 1986). Sampling a motion signal introduces spurious artifacts, whose frequency and amplitude depend on the sampling

rate. Psychophysical measurements show that subjects are able to distinguish sampled from smooth motion if and only if the spurious frequencies produced by the sampling regime are not resolvable, as determined by measuring their thresholds for isolated sinusoids.

The spatio-temporally oriented receptive fields not only allow for the perception of discontinuous motion, but can also cause the image to be interpolated between the positions where it is displayed on each sample. The extrapolation is extremely accurate, and works over long ranges. Indeed, this property can be used to generate complex spatial forms from temporal information alone (Burr & Ross, 1986). When moving forms pass behind a "virtual slatted fence" (allowing information to be displayed only at discrete points), the visual system interpolates between the display points to give the impression of complete spatial forms. Thus, motion detectors not only encode velocity information about moving objects, but also participate in their spatial analysis.

Chromatic and second-order motion

The examples discussed so far refer to motion of objects or images defined by luminance, typically bright or dark lines, sinusoidal gratings or random dot patterns. However, luminance is not the only way to delineate objects: others include color, texture and depth, and all these attributes can support motion. A well-studied example is the equiluminant class of stimuli, defined only by chromatic contrast. Movement of these stimuli yields a sensation of motion, albeit slower and jerkier than that for luminance patterns (Cavanagh, 1991).

Another very common stimulus in recent years is the class defined by variations in contrast, rather than luminance, giving rise to what is now called "second-order" motion (Chubb & Sperling, 1988). A typical example of second-order motion is a field of random dots multiplied (or amplitude-modulated) by a broad moving stimulus, typically a sinusoid. The interesting aspect of this stimulus is that although it gives rise to a strong and compelling sense of motion, neither the Reichardt detector of figure 1 nor the motion-energy detector of figure 2 would respond to it. However, a fairly simple extension can render both models sensitive to second-order motion: all that is needed is a "texture detector", a filter responding to contrast instead of luminance, at the front stage, and the model will respond to amplitude-modulated motion. The "texture detector need not be complicated: a simple half- or full-wave rectifier would suffice. It is still a debated point whether first- and second-order motions are detected by different neural structures, or by essentially the same mechanism with an add-on front-end texture detector. Evidence exists for both possibilities, such as mutual induction of after-effects between the different types of motion, and differential selective activation during fMRI.

Two-dimensional motion

The models shown above are essentially one-dimensional, discriminating leftward from rightward motion. There are various ways of extending these models to cover the two spatial dimensions, such constructing many such units with spatial sub-fields orientated in various directions. Further spatial selectivity can be achieved by extending the spatial filters, or receptive fields, orthogonally to their direction of motion selectivity, emulating the physiological characteristics of receptive fields of mammalian vision. However, these two-dimensional motion units will demonstrate an inherent ambiguity about stimulus direction, usually referred to as the "aperture problem". This stems from the fact that motion along a given trajectory can be decomposed into vectors spanning a range of 180°, so a vast range of detectors will be stimulated by any given trajectory. Various schemes have been proposed for disambiguating the problem, usually involving the combination of signals from more than one detector, either in the form of a "vector sum" of motion units, or "intersection of constraints". There is physiological evidence that the primate visual system adopts one of these schemes (Movshon et al., 1985). When stimulated with "plaids" (two orthogonal sinusoidal gratings) drifting in various directions, neurons in primary visual cortex V1 respond best when the direction of drift is such as to orient one or other of the components appropriately for that neuron, irrespective of the pattern drift. However, in the motion-specialized area MT, neurons respond best when the global motion of the plaid is in the appropriate direction, even though each component is then 45° off-axis. This suggests that as well as being responsible for the opponent stage of the motion detector, MT may help to disambiguate the two-dimensional direction of motion signals.

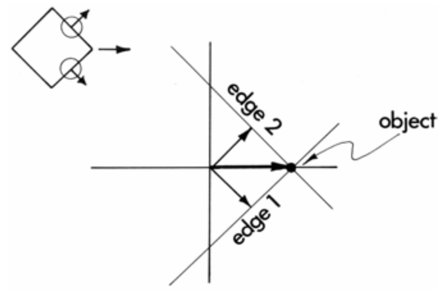

Figure 4 Illustration of the inherent ambiguity of two-dimensional motion. As the diamond moves rightwards, the motion of the edges, within the receptive fields indicated by the circles is diagonally upwards or downwards. In primate cortex, cells in V1 respond to the "component motion" of the edges, while some (but not all) cells in motion area MT respond to the direction of global motion (rightwards).

Other solutions have been proposed for the aperture problem, including the novel suggestion of Bill Geisler (see also Burr, 2000). Geisler points out that given the temporal integration of the visual system, a small, localized target will leave a motion streak, much like the "speed lines" used by cartoonist to caricaturize motion. These static streaks provide potential information to disambiguate direction. A series of masking and motion after-effect studies suggests that this spatial information is in fact integrated with motion information, and may help disambiguation. Another quite different class of experiment has shown that spatial structure of a certain type of moiré pattern can bias otherwise truly apparent motion, showing the influence of static structure on motion direction. Interestingly, however, although the moving streaks may be used to help sense motion, they are not perceived as streaks by the visual system. Although we integrate over time for 120 ms or so, the smear left by moving objects is far less, quite unlike what a camera with that shutter speed would record (Burr & Ross, 1986). Our motion detectors are based on receptive fields that are oriented in space-time, aligning themselves with the motion trajectory, and this should reduce the perceived blur.

This chapter has concentrated on basic motion mechanisms, the early mechanisms that analyze motion locally. Local-motion signals are then combined in various ways, depending on the task. Analysis of optic-flow requires integration of local-motion signals over large areas and complex trajectories. On the other hand, the ability to see transparent motion, and to localize accurately the position of small moving objects requires that the local signals are kept distinct. How these conflicting goals are achieved is very much the subject of much modern research into motion perception.

REFERENCES

Adelson, E.H. & Bergen, J.R. (1985) Spatio-temporal energy models for the perception of motion. J. Opt. Soc. Am., **A2**, 284-299.

Anderson, S.J. & Burr, D.C. (1985) Spatial and temporal selectivity of the human motion detection system. Vision Res., **25**, 1147-1154.

* Burr, D. C. & Ross, J. (1986) Visual processing of motion. Trends in Neuroscience, **9**, 304-306.

* Burr, D.C. (2000). Motion vision: are 'speed lines' used in human visual motion? Current Biol., **10** (12), R440-443.

Burr, D.C. & Ross, J. (1982) Contrast sensitivity at high velocities. Vision Res., **23**, 3567-3569.

Burr, D.C. Ross, J., & Morrone, M.C. (1986) Smooth and sampled motion. Vision Research, **26**, 643-652.

* Cavanagh, P. (1991) Vision at equiluminance. In: J. Cronly-Dillon (Ed.) Visual function and dysfunction: volume 5, chapter 18 (pp. 234-250). London: Macmillan.

Chubb, C. & Sperling, G. (1988) Drift-balanced random stimuli: a general basis for studying non-Fourier motion perception. J. Opt. Soc. Am, **A 5**, 1986-2007.

Geisler, W.S. (1999) Motion streaks provide a spatial code for motion direction. Nature, **400**, 65-69.

Heeger, D.J., Boynton, G.M., Demb, J.B., Seidemann, E. & Newsome, W.T. (1999) Motion opponency in visual cortex. J. Neuroscience, **19**, 7162-7174.

Movshon, J.A., Adelson, E.H., Gizzi, M.S. & Newsome, W.T. (1985). The analysis of moving visual patterns. In: R.G. C. Chagas, C. Gross (Ed.) Pattern Recognition Mechanisms, 54 (pp. 117-151): Pontificiae Academiae Scientiarum Scripta Varia.

Qian, N. & Andersen, R. (1994) Transparent motion perception as detection of unbalanced motion signals. II. Physiology. J. Neurosci, **14**, 7367-7380.

Reichardt, W. (1961) Autocorrelation, a principle for evaluation of sensory information by the central nervous system. In: W. Rosenblith (Ed.) Sensory communications (New York: John Wiley.