

Neural dynamics of motion segmentation

Stephen Grossberg*

Ennio Mingolla†

Center for Adaptive Systems
Boston University,
111 Cummington Street
Boston, MA 02215 (USA)

*Supported in part by AFOSR F49620-86-C-037,
AFOSR F49620-87-C-0018, ARO DAAL03-88-K-0088

†Supported in part by AFOSR F49620-87-C-0018

Abstract

A neural network model of motion segmentation by visual cortex is described. The model clarifies how preprocessing of motion signals by a Motion Oriented Contrast Filter (MOC Filter) is joined to long-range cooperative motion mechanisms in a motion Cooperative Competitive Loop (CC Loop) to control phenomena such as induced motion, motion capture, and motion aftereffects. The total model system is a motion Boundary Contour System (BCS) that is computed in parallel with the static BCS of Grossberg and Mingolla (1985a, 1985b, 1987) before both systems cooperate to generate a boundary representation for three dimensional (3-D) visual form perception. The present investigations clarify how the static BCS can be modified for use in motion segmentation problems, notably for analysing how ambiguous local movements (the aperture problem) on a complex moving shape are suppressed and actively reorganized into a coherent global motion signal. Unlike many previous approaches, we analyse how a coherent motion signal is imparted to all regions of moving figure ("motion capture"), not only regions at which unambiguous motion signals exist.

Keywords: motion, segmentation, grouping, cooperative, competitive, filtering

1. Introduction: Why are Static and Motion Boundary Contour Systems Needed?

Some regions of visual cortex are specialized for motion processing, notably region MT (Albright, Desimone, & Gross, 1984; Maunsell & van Essen, 1983; Newsome, Gizzi, & Movshon, 1983; Zeki, 1974a, 1974b). However, even the earliest stages of visual cortex processing, such as simple cells in V1, require stimuli that change through time for their maximal activation and are direction-sensitive (DeValois, Albrecht, & Thorell, 1982; Heggelund, 1981, Hubel & Wiesel, 1962, 1968, 1977; Tanaka, Lee, & Creutzfeldt, 1983). Why has evolution generated regions such as MT, when even V1 is change-sensitive and direction-sensitive? What computational properties are achieved by MT that are not already available in V1?

The monocular Boundary Contour System (BCS) theory of Grossberg and Mingolla (1985a, 1985b, 1987), and its binocular generalization (Grossberg, 1987a, Grossberg & Marshall, 1989), has successfully modelled many boundary segmentation properties of V1 and its prestriate projections. (See Grossberg, 1987b, 1988 for summaries of these and related articles.) The BCS has until now been used to analyse data generated in response to static visual images. Henceforth we therefore call such a BCS a static BCS model. Nonetheless its model cells can easily be gated by cells sensitive to image transients, such as Y cells (Enroth-Cugell & Robson, 1966; Hoffman, 1973; Sekuler, 1975, Stone, 1972, Stone & Dreher, 1973; Tolhurst, 1973) to generate receptive fields sensitive to image transients. How does a motion BCS differ from a static BCS whose cells are sensitive to image transients?

2. Static and Motion Filtering: Direction-of-Contrast and Direction-of-Motion

That boundaries of opposite direction-of-contrast are perceptually linked is vividly illustrated by the reverse-contrast Kanizsa square. (See Figure 1.) A fundamental property of the front end of the BCS, which is a Static Oriented Contrast Filter (SOC Filter), is

that its output is *insensitive to direction-of-contrast*, in order to support perception of boundaries in variable illumination. Figure 2 illustrates how this insensitivity is achieved through the pooling by units identified with complex cells of rectified outputs of units identified with simple cells, whose receptive fields are elongated and sensitive to opposite contrast polarities. This pooling operation implies that *the complex cell layer of the SOC Filter is insensitive to direction-of-motion*, as well as to direction-of-contrast, because of the simple-to-complex pooling shown in Figure 2. Evidently, any useful filter that will act as the front-end of a motion segmentation system must be *sensitive to direction-of-motion* while being *insensitive to direction-of-contrast*.

3. Global segmentation and grouping: From locally ambiguous motion signals to coherent object motion signals

In their discussion of "velocity space," Adelson and Movshon (1980, 1982) introduce diagrams similar to Figure 3a to illustrate local motion

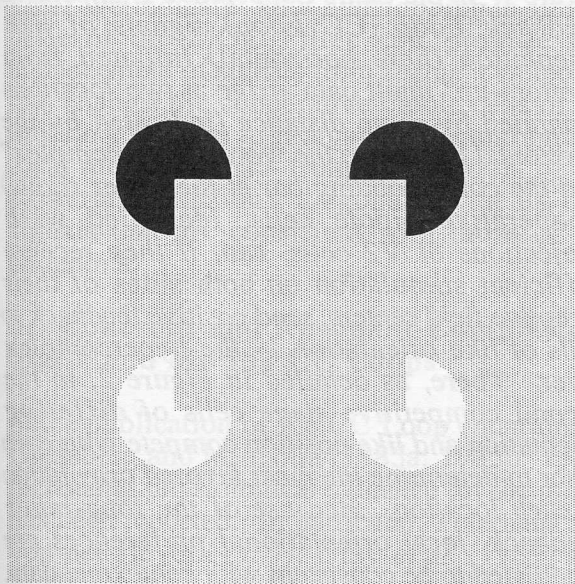


Figure 1: In the reverse-contrast Kanizsa square a perceptual completion occurs between dark-to-light and light-to-dark vertical edges, illustrating the complementarity of visual representation of boundaries (involved in segmentation), which are insensitive to direction-of-contrast and features (brightness), which are sensitive to direction-of-contrast.

direction (and speed) ambiguity from information confined to an aperture. In this half of the figure, the length of arrows codes possible trajectories of the point A which would be consistent with the measured change of contrast over time of the cell in question; for this reason, it is sometimes said that a cell such as that whose response is illustrated is sensitive to only the normal component of velocity. Another way of thinking about this situation is illustrated in Figure 3b, in which the length of arrows is roughly proportional to the cell's "prior probability distribution" for interpreting changing stimulation as occurring in one of several directions, of which the direction perpendicular to the cell receptive field's axis of orientation is locally preferred. Note that in

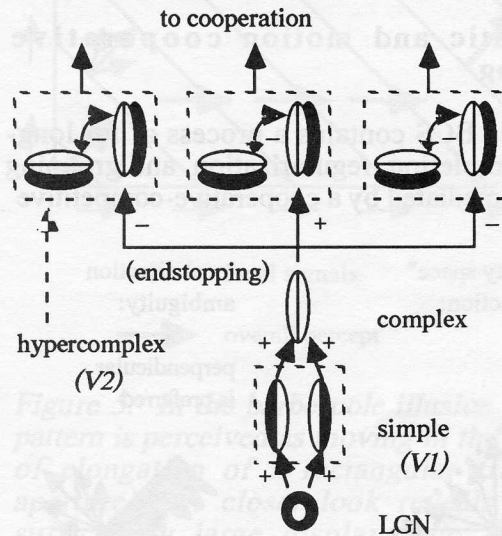


Figure 2: Stages of the Static Oriented Contrast (SOC) Filter: Simple cell receptive fields are oriented and sensitive to direction-of-contrast. Rectified outputs from pairs of simple cells sensitive to opposite directions-of-contrast input to complex cells, which are sensitive to amount-of-contrast, but not to direction-of-contrast or to direction-of-motion. Complex cells input to two successive stages of hypercomplex cells. At the first competitive stage a complex cell excites like-oriented hypercomplex cells at its position and inhibits, via an endstopping operation, hypercomplex cells at nearby positions. At the second competitive stage, hypercomplex cells sensitive to different orientations compete with one another. Inhibition is maximal between perpendicular orientations.

this conception, if a cell with an oriented receptive field (e.g. a simple cell) is being stimulated by an edge that is not perfectly aligned with its receptive field's dark-to-light contrast axis, it's "preferred direction" will not correspond to that perpendicular to the edge. In this case, however, it is assumed that within a hypercolumn of cells tuned to similar spatial frequency, contrast, and temporal parameters but varying in preferred orientation, some other cell whose preferred orientation was more nearly aligned with the edge would generate a stronger signal than the cell in question. Thus, the distribution of motion signals across cells tuned to all orientations would favor the direction perpendicular to the orientation of the edge.

4. Static and motion cooperative grouping

The static BCS contains a process of for long-range completion, regularization, and grouping which is mediated by a cooperative-competitive

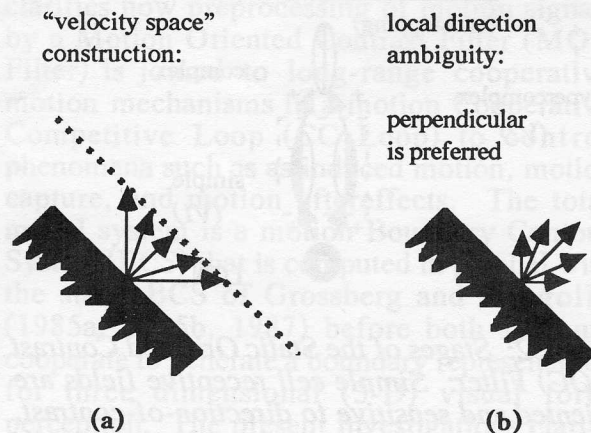


Figure 3: (a) This "velocity space" construction, adapted from Adelson and Movshon (1980, 1982), illustrates that, for a given measurement of image change by a localized receptive field, the motion of a given point on a moving edge could be along any of the trajectories whose arrows end on the dashed line. (b) A complementary way to view this direction ambiguity (aperture problem) is to consider that local motion signals originating from an edge are strongest in the (locally preferred) perpendicular direction, while covering a range of possible directions.

feedback loop (CC Loop) whose competitive layer is identified with hypercomplex cells of V2 and whose cooperative layer contains units called "bipole cells," which are hypothesized to exist in the projections of V2 cells. The CC Loop seeks to form and sharpen boundaries whenever evidence from bottom-up inputs in

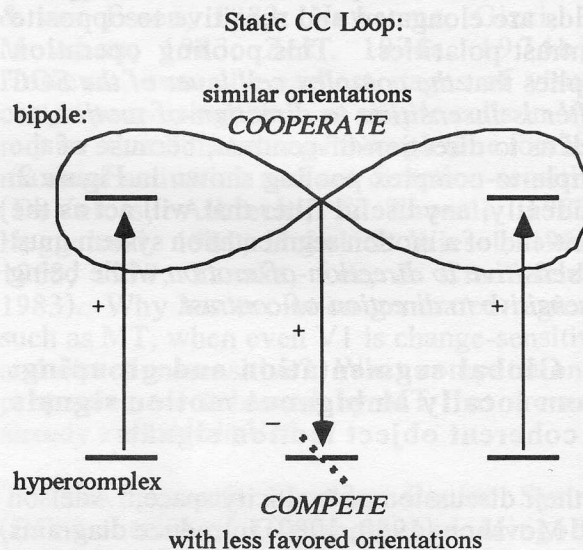


Figure 4: In the static Cooperative-Competitive feedback Loop (CC Loop) signals of like orientation (here horizontal) initiated from "bottom-up" input data arrive at two locations separated by some distance (indicated by the horizontal bar above the word "hypercomplex" and the corresponding bar at the lower right.) Cooperative bipole cells, identified with projections of V2 cells, can, if they receive sufficient stimulation on both sides of their receptive field centers, send positive feedback to cells of like orientation in the hypercomplex layer. There, as detailed in Figure 2, in the second competitive stage cells of differing orientation and like position compete with each other to determine the most favored orientation at each location. Because of the long-range feedback, local orientational preferences are modulated to be consistent with contextual data. While in this example the completion is colinear, the "figure 8" shape of the bipole cell indicates that completion can be curvilinear, so long as a consistent cooccurrence of oriented signals favors interpolation at the bipole center. See Figure 31 and equations A13 through A21 of the Appendix of Grossberg & Mingolla, (1985) for a detailed explanation of static bipole cells.

two regions indicates that a colinear (possible curved) continuation of boundary activity is called for. (See Figure 4.) A horizontally tuned bipole cell sends feedback to horizontally tuned cells in the competitive layer.

In considering how the static CC Loop must be modified to deal with motion segmentation, consider that motion is not binary but continuously valued; headings can be, for example, "north by northwest." The analysis of moving contours thus requires an *additional degree of freedom* than the analysis of static contours, for a contour of a given orientation can be moving in an infinity of directions, and conversely contours of an infinity of orientations can be moving in the same direction; thus a modification in the structure of the static BCS is required. Consider again the aperture problem (Wallach, 1976). In the barberpole illusion, the perception of motion direction along entire contours -- whose measurement by cells with localized receptive fields is everywhere subject to the aperture problem -- is determined by the perceived motion of their endpoints. (See Figure 5.) Later we indicate how *endstopping* in simple cells of the MOC Filter can provide the enhancement of signals from segment endpoints that enables the cooperative bipole cells of the motion CC Loop to reorganize the ambiguous local motion signals from the interiors of the barberpole illusion's diagonal segments into signals that are consistent with those of the endpoints.

5. Coherent and Resonant Completion: Static and Motion CC Loops

Since publication of the CC Loop equations in Grossberg and Mingolla (1985a; 1985b; 1987) increasing physiological evidence for the model has been reported. In particular, Peterhans and von der Heydt (1989), Gray *et al.* (1989) and Eckhorn *et al.* (1989) have recently reported evidence for cooperative interactions in V2. Given that the static CC Loop possesses a powerful mechanism for enhancing consistent signals and suppressing inconsistent or ambiguous signals from orientationally-tuned cells, and given that *in vivo* many early orientationally tuned cells, including simple and complex cells, already show some sensitivity to motion, the question naturally arises whether

necessary motion segmentation and grouping operations could be accomplished by the previously described (static) CC Loop. Strong indications that such could not be the case arise from considering the most basic aspects of the dynamic geometry of form perception for static and moving contours. A *static* form system is concerned with *orientation* of contours, while a *motion* form system is concerned with *direction* of moving contours. Within the space of orientations, opposites of given signals occur at the perpendiculars to those signals (90°); within the space of directions; opposites occur at straight angles (180°). Thus keeping track of direction as well as orientation requires an additional degree of freedom, for as we have

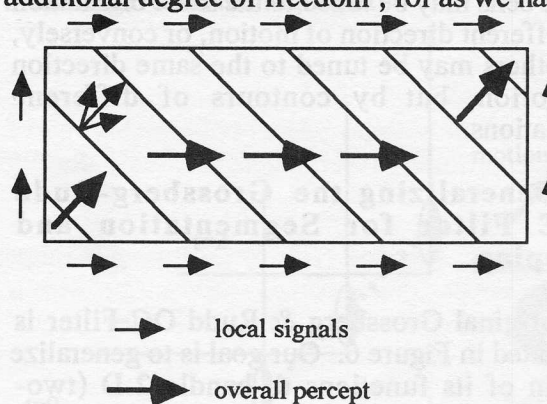


Figure 5: In the barberpole illusion a striped pattern is perceived as moving in the direction of elongation of a rectangular frame (or aperture). A closer look reveals that for sufficiently large displays the perceived direction of motion varies within the display. The perceived motion near the lower left and upper right corners of the rectangle is diagonal, while horizontal motion is seen through the bulk of the display. Local motion signals, indicated by thin arrows at only one location within the rectangle, are ambiguous throughout the interior lengths of each diagonal line. Unambiguous signals are generated where each diagonal line meets the horizontal or vertical contours of the rectangle. (These signals are diagramed outside the rectangle for clarity.) Evidently, such signals exert an influence on the percept that is disproportionate to their areal extent, since the overall percept throughout a diagonal line tends to be a resultant (horizontal plus vertical to diagonal or horizontal plus horizontal to horizontal) of those signals.

noted, orientation of contrast and direction of motion are separate dimensions, each requiring specialized machinery for processing -- which is another way of viewing the aperture problem. Note that the preceding taxonomy of "static" and "motion" CC Loops is heuristic and not to be taken literally; in particular, it is not meant to imply logical exclusivity of function. Thus, when a contour moves, the *static* CC Loop may continue to operate to best determine coherent *orientation* of the moving contour (within certain parameters of spatiotemporal resolution) at the same time that the *motion* CC Loop determines the *direction* of motion of that contour. In the motion CC Loop, then, two bipole cells may be tuned to the same orientation but different direction of motion, or conversely, two others may be tuned to the same direction of motion, but by contours of different orientations.

6. Generalizing the Grossberg-Rudd MOC Filter for Segmentation and Grouping

The original Grossberg & Rudd OC Filter is illustrated in Figure 6. Our goal is to generalize certain of its functions to handle 2-D (two-dimensional) motion segmentation issues.

The full domain of motion segmentation and grouping includes such problems as determining structure in depth from motion, motion transparency, and motion grouping amid occlusion. Although the motion BCS is conceived with these and related difficult phenomena in mind, we will not attempt to deal with the full range of issues these phenomena raise at present, but will instead focus on the elementary grouping operations necessary to perform essentially 2-D detections of object motion within the visual field. Even here difficult issues arise, as we illustrate with a simple example. Consider the lower right corner of a homogeneous rectangular form of relatively high luminance that is moving diagonally upward and to the right on a homogeneous background of relatively low luminance. (See Figure 7.) Both the regions of horizontal and vertical contrast near the corner provide signals to the MOC Filter, provided that the sustained cells of Level 2 (Figure 6) are taken to be spatially laid out as indicated in Figure 8. (Of course, motion signals of many

other directions will also be generated along the lengths of the horizontal and vertical edges; these will be considered subsequently.) In other words, for at least some of the gating nodes of Layer 4 (Figure 6), the layout of receptive field centers of contributing sustained cells of Layer 2 is taken to be in a direction *diagonal* to the orientational preference of the individual sustained cells. It would make no sense to build a motion filter whose receptive field centers were arrayed *colinearly* with the contributing sustained cell's orientational preference -- although this type of arrangement

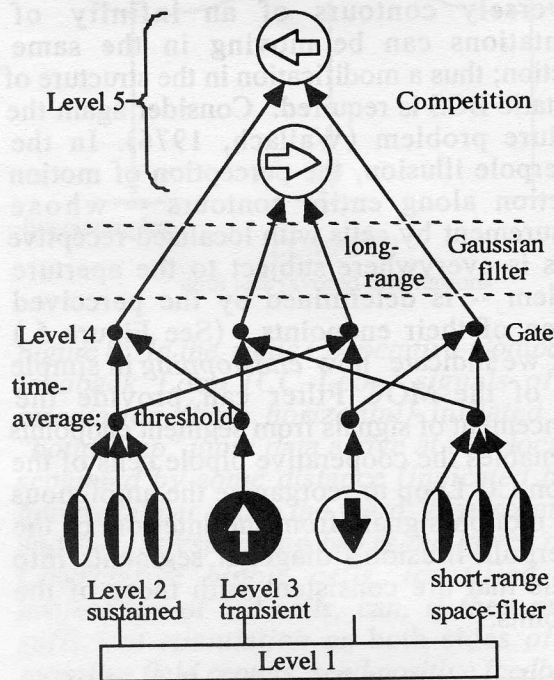


Figure 6: The Motion Oriented Contrast (MOC) Filter is insensitive to direction-of-contrast but sensitive to direction-of-motion. Level 1 registers the input pattern. Level 2 consists of sustained response cells with oriented receptive fields that are sensitive to direction-of-contrast. Level 3 consists of transient response cells with unoriented receptive fields that are sensitive to direction of change in the total cell input. Level 4 cells combine sustained cell and transient cell signals to become sensitive to direction-of-motion and sensitive to direction-of-contrast. Level 5 cells combine Level 4 cells to become sensitive to direction-of-motion and insensitive to direction-of-contrast.

might be suitable for colinear completion in a static form system. Accordingly, we propose that a variant of a "sine law" exists, whereby the contribution of any sustained cell at Level 2 to Level 4 gating cell is modulated by the (absolute value of the) sine of the angle formed between the sustained cell's *orientational preference* and the gating cell's *directional preference*.

The long range filter (Level 5, Figure 6) can simultaneously accept motion signals from both the horizontal and vertical edges of the moving corner, *despite the gating of one set of signals by transient "luminance increasing" detectors (Level 3, Figure 6) and gating of another set by "luminance decreasing" detectors*. Thus while simultaneous increase and decrease of luminance is logically impossible in an infinitesimal area, and while a too rapid change from increase to decrease may be unresolvable by sustained cells at Level 2, the simultaneous *nearby increase and decrease of luminance with a coherent trajectory or direction despite different contour orientations* is fodder for the long-range filter. Note that the long-range filter of the MOC Filter is *not the same* as the long-range grouping stage of cooperative bipole cells.

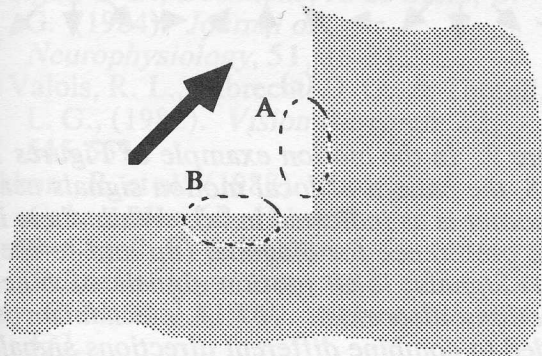


Figure 7: The lower right corner of a horizontally oriented rectangular region of homogeneous high luminance moves diagonally upward and to the right over a background of homogeneous low luminance. In region A a dark-to-light (luminance increasing over time) transition occurs at a vertical edge, while in region B a light-to-dark (luminance decreasing over time) transition occurs at a horizontal edge.

7. Endstopping: Generation of a terminator or corner advantage in motion signals

In discussing the barberpole illusion we referred to an "advantage" for motion signals near terminators or corners of contours. (See Figure 3.) The designation "advantage" connotes that those signals tend to be better indicators of object motion than signals generated from a relatively straight interior of a contour. For this advantage to be manifest in perception, however, that advantage must also be one of signal *strength*, the more so because the regions or spatial extent of interior motion signals is often larger than the region of

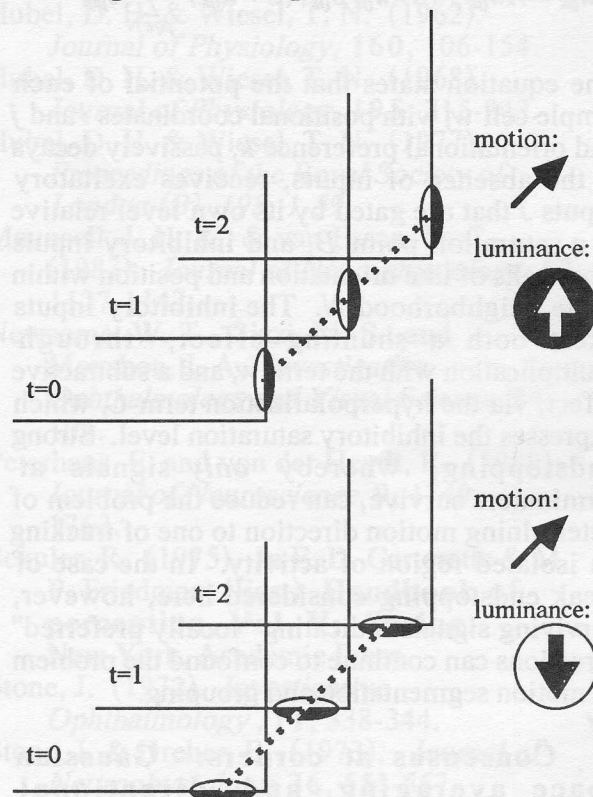


Figure 8: Over three successive time increments, the contours of the rectangle of Figure 7 occur in the positions indicated, while luminance increases along the vertical edge and decreases along the horizontal edge. If certain of the sustained cell receptive fields sending inputs to Level 4 of the MOC Filter (Figure 6) were arranged as indicated, a diagonal motion signal could be generated from both vertically and horizontally oriented cells, in conjunction with luminance gating signals of opposite signs.

terminator or corner signals. The source of the advantage would appear to involve *endstopping* at the very front end of the MOC Filter. Many simple cells, identified with the orientation and direction-of-contrast sensitive sustained cells of Level 2 of the MOC Filter, exhibit endstopping (Dreher, 1972.) (Note that this endstopping is functionally analogous to the endstopping of the first competitive stage of the SOC Filter (Figure 2.)

Endstopping may be described by a shunting competitive network equation with hyperpolarization:

$$\frac{d w_{ijk}}{dt} = -A w_{ijk} + (B - w_{ijk}) J_{ijk} - (C + w_{ijk}) \sum_{pq \in N} J_{pqk}$$

The equation states that the potential of each simple cell w , with positional coordinates i and j and orientational preference k , passively decays in the absence of inputs, receives excitatory inputs J that are gated by its own level relative to a saturation point B , and inhibitory inputs from cells of like orientation and position within some neighborhood N . The inhibitory inputs exert both a shunting effect, through multiplication with the term w , and a subtractive effect, via the hyperpolarization term C , which expresses the inhibitory saturation level. Strong endstopping, whereby *only* signals at terminators survive, can reduce the problem of determining motion direction to one of tracking an isolated region of activity. In the case of weak endstopping considered here, however, surviving signals indicating "locally preferred" directions can continue to confound the problem of motion segmentation and grouping.

8. Consensus at corners: Gaussian space averaging and directional competition

In the weak endstopping case the local motion signals from the lower right corner of our moving rectangle would have roughly the form diagrammed in Figure 9a. Note that the domain of spatial averaging of the Gaussian filter (transition from Level 4 to Level 5 of the MOC Filter) is presumed to be large enough to span the signals generated by the ends of the leading vertical and trailing horizontal edges. At Level 5, then, signals of many directions occur for cells coding the same position. Those directions will have the appropriate "central

tendency", however, and a simple center-surround competition in the space of directions, analogous to the revised version of the second

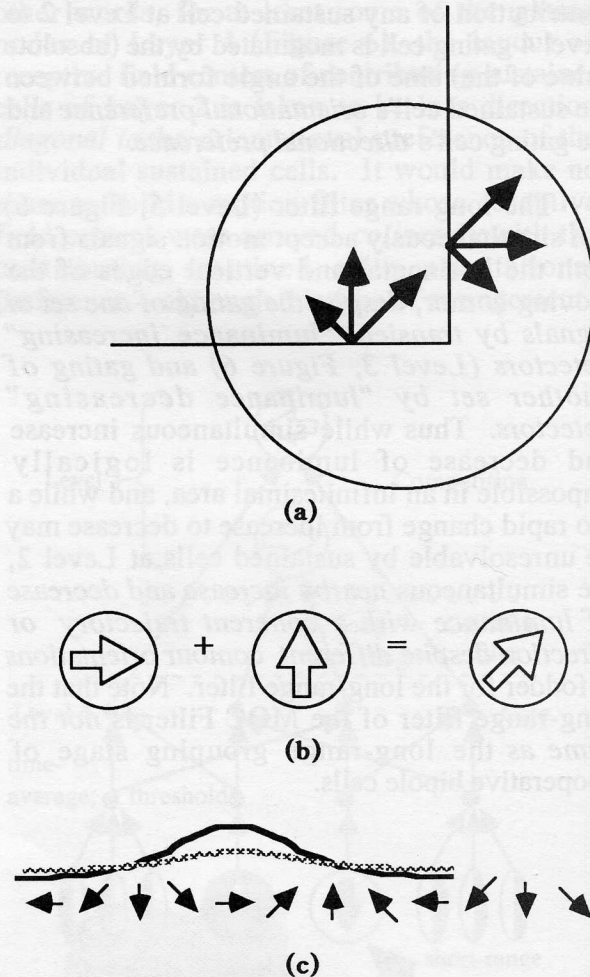


Figure 9: In the motion example of Figures 7 and 8, the pattern of local motion signals near the corner is as indicated in (a). While there is some preference for diagonal (up-and-to-the-right) signals, local motion signals of other directions also exist. (b) A mechanism is needed to combine different directions signals into a single coherent local direction signal. (c) The signal combination can be accomplished by a motion analog of the second competitive stage among orientations of the SOC Filter (Figure 2), as described in Grossberg & Mingolla, (1987). A excitatory on-center, inhibitory off-surround network organization among cells coding different directions-of-motion at the same position can accomplish the desired pooling and choice through competitive peak summation and sharpening.

competitive stage (for orientations) of the static BCS described by Grossberg & Mingolla (1987), suffices to choose the direction which is most consistent with surrounding input data at each location. (See Figure 9b.)

9 Contextually consistent and coherent motion segmentation

In this article we have described motion analysis mechanisms whereby the visual system frees itself from an excessive reliance on either purely local (short-range filtering) computations or top-down (cognitive or expectancy based) computations. Instead, within a *perceptual* middle ground, competitive and cooperative interactions within a parallel and structured network with several scales of interaction help to choose and enhance those aspects of local data which contribute to coherent and consistent measures of object motion.

References

- Adelson, E. H. & Movshon, J. A. (1980). *Journal of the Optical Society of America*, **70**, 1605.
- Adelson, T. & Movshon, J. A. (1982). *Nature*, **300**, 523-525.
- Albright, T. D., Desimone, R. & Gross, C. G. (1984). *Journal of Neurophysiology*, **51**, 16-31.
- De Valois, R. L., Albrecht, D. G., & Thorell, L. G., (1982). *Vision Research*, **22**, 545-559.
- Eckhorn, R. et al. (1988). *Biological Cybernetics*, **60**, 121-130.
- Enroth-Cugell, C. & Robson, J. G. (1966) *Journal of Physiology*, **187**, 517-552.
- Gray, C. M. et al. (1989). *Nature*, **338**, 334-337.
- Grossberg, S. (1987a) *Perception and Psychophysics*, **41**, 87-116.
- Grossberg, S. (Ed.) (1987b). **The adaptive brain, Vol. II: Vision, Speech, Language, and Motor Control.** Amsterdam: North-Holland.
- Grossberg, S. (Ed.) (1988) **Neural Networks and natural intelligence.** Cambridge, MA: MIT Press.
- Grossberg, S. & Marshall, J. (1989). *Neural Networks*, **2**, 29-51.
- Grossberg, S. & Mingolla, E. (1985a). *Psychological Review*, **92**, 173-211.
- Grossberg, S. & Mingolla, E. (1985b). *Perception and Psychophysics*, **38**, 141-171.
- Grossberg, S. & Mingolla, E. (1987). *Computer Vision, Graphics, and Image Processing*, **37**, 116-165.
- Grossberg, S. & Rudd, M. (1989). A neural architecture for visual motion perception: Group and element apparent motion. *Neural Networks*, **2**, 421-450.
- Heggelund, P. (1981). *Experimental Brain Research*, **42**, 89-98.
- Hoffman, K. P. (1973). *Journal of Neurophysiology*, **36**, 409-424.
- Hubel, D. H. & Wiesel, T. N. (1962). *Journal of Physiology*, **160**, 106-154.
- Hubel, D. H. & Wiesel, T. N. (1968). *Journal of Physiology*, **195**, 215-243.
- Hubel, D. H. & Wiesel, T. N. (1977). *Proceedings of the Royal Society of London (B)*, **198**, 1-59.
- Maunsell, J. H. R., & van Essen, D. C. (1983). *Journal of Neurophysiology*, **49**, 1127-1147.
- Newsome, W. T., Gizzi, M. S., and Movshon, J. A. *Investigative Ophthalmology and Visual Science*, **24**, 106.
- Peterhans, E. and von der Heydt, R. (1989). *Journal of Neuroscience*, **9**, 1749-1763.
- Sekuler, R. (1975). In E. D. Carterette & M. P. Friedman, (Eds.) **Handbook of perception, Vol. V: Seeing.** New York, Academic Press.
- Stone, J. (1972). *Investigative Ophthalmology*, **11**, 338-344.
- Stone, J. & Dreher, B. (1973). *Journal of Neurophysiology*, **36**, 551-567.
- Tanaka, M., Lee, B. B., and Creutzfeldt, O. D. (1983). In J. D. Mollon and L. T. Sharpe (Eds.) **Colour Vision.** New York, Academic Press.
- Tolhurst, D. J. (1973). *Journal of Physiology*, **231**, 385-402.
- Wallach, H. (1976). **On perception.** New York, Quadrangle.
- Zeki, S. M. (1974a). *Journal of Physiology (London)*, **236**, 549-573.
- Zeki, S. M. (1974b). *Journal of Physiology (London)*, **242**, 827-841.