The Role of Terminators and Occlusion Cues in Motion Integration and Segmentation: A Neural Network Model

Lars Lidén[1] Christopher Pack[2*]

[1]Department of Cognitive and Neural Systems
Boston University
677 Beacon St.
Boston, MA 02215

[2]Department of Neurobiology
Harvard Medical School
220 Longwood Ave.
Boston, MA 02115

*To whom correspondence should be addressed
email cpack@hms.harvard.edu
FAX: 617-734-7557

Key Words: motion integration, motion segmentation, MT, terminators, occlusion, modeling

Running head: Motion integration and segmentation

February, 1999

Acknowledgments - The authors wish to thank two anonymous reviewers for helpful comments on a previous version of this manuscript. Lars Lidén was supported by The Krell Institute's Computational Science Graduate Fellowship Program. Christopher Pack was supported by grants from the Office of Naval Research (ONR N00014-94-1-0597, ONR N00014-95-1-0409, ONR N00014-95-1-0657). A preliminary report has appeared previously (Lidén, 1998).

Abstract

The perceptual interaction of terminators and occlusion cues with the functional processes of motion integration and segmentation is examined using a computational model. Integration is necessary to overcome noise and the inherent ambiguity in locally measured motion direction (the aperture problem). Segmentation is required to detect the presence of motion discontinuities and to prevent spurious integration of motion signals between objects with different trajectories. Terminators are used for motion disambiguation, while occlusion cues are used to suppress motion noise at points where objects intersect. The model illustrates how competitive and cooperative interactions among cells carrying out these functions can account for a number of perceptual effects, including the chopsticks illusion and the occluded diamond illusion. Possible links to the neurophysiology of the middle temporal visual area (MT) are suggested.

Introduction

The processing of motion is perhaps the most fundamental visual task of any biological system. When an object in the environment moves, an observer must be able to estimate its trajectory in three-dimensions in order to initiate or avoid contact. Physiological and theoretical work suggests that the extraction of motion information takes place in at least two stages (Adelson & Movshon, 1982, Yuille & Grzywacz, 1988). The first stage extracts local motion measurements, which typically correspond to a small part of an object in motion. The second stage combines local motion measurements to form a globally coherent motion percept.

There are several problems inherent to the detection of local motion that invariably result in inaccurate and non-unique motion measurements. In general, motion computation falls into the category of problems which are ill-posed (Poggio & Koch 1985). A key problem for local motion estimation is what is known as the *aperture problem* (Marr & Ullman 1981), which states that any localized motion sensor can only detect motion orthogonal to a local contour. Such motion measurements are *ambiguous*, in the sense that any direction within a 180° range is equally compatible with the local motion measurement (Figure 1). A local motion detector, therefore, cannot generally compute the direction of motion of an object that is bigger than its own field of view. Since early detection of motion is carried out by neurons with relatively small receptive fields, the aperture problem is an immediate difficulty for biological visual systems. In contrast to motion measurements along a contour, motion measurements at the termination of a contour are *unambiguous*, since only one direction of motion is compatible with the motion of a con-

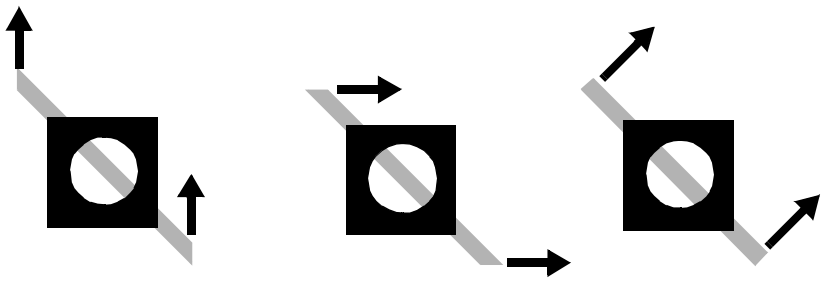

Figure 1: The Aperture Problem: Depicted above are three lines moving in different directions beneath square occluders with small holes (apertures). When viewed through the apertures, the three motions appear identical.

tour's terminators. An effective motion processing system must use unambiguous motion signals in one region of the image to constrain the interpretations of ambiguous local motion directions in other regions.

In order to address the problems inherent to the first stage of motion computation, it is advantageous to subdivide the second stage into two parallel computations, one for *integration* and the other for *segmentation* (Braddick, 1993). Integration is a process that combines noisy local motion signals in such a way that spurious estimates are averaged out, and the local aperture problem is overcome. Segmentation detects the presence of motion discontinuities and prevents integration of motion signals between objects with different trajectories. Physiological (Allman & McGuinness, 1985; Tanaka, Hikosaka, Saito, Yukie, Fukada & Iwai, 1986; Born & Tootell, 1992; Saito, 1993) and psychological studies (Vaina & Grzywacz, 1992; Vaina, Grzywacz & Kinkinis, 1994) support the idea that global motion processing is subdivided into these two computational processes.

Psychophysical evidence suggests that the integration of motion signals takes time (hundreds of milliseconds) and can have effects across the entire visual field (Williams &

Sekuler, 1984; Ramachandran & Anstis, 1986a; Watanabe & Cole 1995). While the detection of local motion signals can be viewed as essentially instantaneous, the global motion percept develops as a consequence of interactions across time and space. These interactions are likely to consist of excitation and inhibition which can be detected at the level of individual neurons (Snowden and Braddick, 1989), but propagate across large regions of visual space (Watanabe and Cole, 1995).

An additional factor to consider is the influence of static surfaces on motion perception. An array of psychophysical data suggests that surface perception profoundly alters the integration of motion signals (Stoner & Albright, 1993). Information from object shape (Tse, Cavanagh & Nakayama, 1998) and static depth cues (Ramachandran and Anstis, 1986b; Shimojo, Silverman, & Nakayama, 1989) can strongly influence the perception of motion. In particular, static cues are extremely useful for distinguishing between intrinsic terminators, which signal the physical end of an edge of an object, and extrinsic terminators, which are caused by occlusion. Intrinsic terminators provide an unambiguous motion signal regarding the true direction of object motion, while extrinsic terminators provide a locally ambiguous signal which must be suppressed for accurate motion computation (Shimojo, Nakayama, & Silverman, 1989). The perception of occlusion is therefore crucial to the processing of motion (Castet et al., in press; Lidén, 1998). An important cue for occlusion is the presence of a T-junction where object edges intersect (Cavanagh, 1987).

This paper describes a model which demonstrates how perceptual phenomena of motion propagation over time can be related to neural interactions which gradually propa-

gate motion signals across cells representing different regions of space. Global percepts emerge as a result of excitatory and inhibitory interactions among cells encoding motion direction in local regions of space. The model illustrates how the two computational requirements of integration and segmentation of motion signals can be implemented by a single motion processing system, and how these systems can interact with static form information concerning occlusion. Interactions between integration and segmentation processes allow the system to eliminate noise and overcome the aperture problem, while preserving useful information at motion discontinuities. A series of model simulations shows that many perceptual phenomena can be explained in terms of the interactions between the proposed motion processing subsystems.

A Neural Network Model

The model consists of a neural network with two primary systems, each composed of interconnected cells. Model cells respond preferentially to stimuli moving in a particular direction within a specific region of space, known as the receptive field center. A concentric region outside the receptive field center is called the surround, by analogy with center-surround cells found in motion processing areas such as middle temporal (MT) cortex. The first system comprises a set of *integration cells*, which respond best when motion in both the center and surround are in the same direction. This type of cell performs a smoothing or averaging process which is used to overcome noise and the aperture problem by propagating motion signals across visual space.

The second system comprises a set of *segmentation cell*s which possess a surround

that is active when stimuli move opposite to the directional preference of the center. The surround also inhibits cell activity when stimuli move in the center's preferred direction. As a result, segmentation cells do not respond well to large fields of coherent motion. Instead, this type of cell signals the presence of motion discontinuities and is used to constrain the interactions of the integration cells.

The recurrent connectivity between model cells is crucial to the processing of motion signals. Cooperative interactions between integration cells at different spatial locations enhance the activity of disambiguated motion signals and propagate these signals across visual space. Although the sizes of the receptive fields are relatively small with respect to the size of the image, disambiguating signals travel across space through connections among nearby nodes. The interaction between segmentation and integration cells is also recurrent, and equally important. The spread of signal from integration cells is constrained by the activity of segmentation cells, which signal motion boundaries. However, the activity of segmentation cells depends critically on input from integration cells. The dynamics of these interactions determines the evolution of the global motion percept.

Input Nodes (LMDs)

The model requires as input an analog value corresponding to the evidence for motion in each direction. Since the details of the initial motion extraction mechanism are of little importance in this context, a simple correlation scheme (e.g., Reichardt, 1961; Van Santen & Sperling, 1984, 1985) was chosen for it computational simplicity. More complex energy models (Fennema & Thompson, 1979; Marr & Ullman, 1981; Adelson & Bergen,

1985; Grossberg & Rudd, 1989, 1992) could also serve as a front-end to the current model. For the simulations described below, two successive image frames were used to compute correlation. For each position in space, a small window of pixels was chosen from the first frame. The grey-level pixel values for the window in the first frame were compared to the grey-level values for shifted windows in the second frame. The correlation between the two grey-levels was used as a measurement of the motion in the direction of the shift. For each position eight shift directions were employed. For each direction, shifts of one and two pixels were measured and the resulting correlation values summed. The use of multiple shift sizes was employed to capture a larger range of speeds. Details of these calculations can be found in Appendix 1. In the following model description, the input cells are referred to as local motion detectors (LMDs).

Integration (*I*) Cells

The dynamics of the integration (*I*) cells depend on three computational principles (See Figure 2):

[1] A *directionally dependent inhibition* exists between cells with different direction tuning at a given spatial location weighted by the directional difference. In this way, cells in ambiguous locations (areas containing multiple active directions) are preferentially suppressed with respect to cells in unambiguous locations (areas with one or few active directions). It has been suggested that such preferential weighting of unambiguous motion signals is used by the visual system to overcome the aperture problem (Shiffrar & Pavel, 1991; Lorenceau & Shiffrar, 1992).

[2] A *directionally dependent excitation across space*, weighted by the magnitude of directional difference, operates between nearby cells. In the model this excitation signal is used to propagate disambiguated motion signals across space. Motion signals that are propagated to regions lacking direct LMD input will be referred to as *sub-threshold* inputs, since the allowed maximal activation level of these nodes is less than when LMD input is present. This sub-threshold activity allows disambiguating motion signals from one area of space to influence motion in another even when the motion signals are spatially disconnected.

[3] A *directionally dependent long range inhibition*, weighted by the magnitude of directional difference, operates between cells of different directional tuning. This inhibition indirectly implements the constraint propagation proposed by Marr (1982), by suppressing integration cell activity across neighboring regions containing conflicting motion estimates. In this way the model can still achieve a rudimentary segmentation even when only the *I* cells are included (Lidén 1997). The importance of long range inhibition has been demonstrated for the early stages of motion processing in both the retina (Barlow & Levick, 1965) and the primary visual cortex (Sillito, 1975), and in other modeling work (Qian, Andersen & Adelson, 1994; Grunewald, 1996; Chey et al., 1997). Details of computations for the integration cells can be found in Appendix 2.

Segmentation (*S*) Cells

The second model system is a segmentation system, consisting of cells with inhibitory surrounds. For each position in visual space there is a set of segmentation (S) cells tuned for the full set of possible directions of motion (in this case 8 directions are used). There are

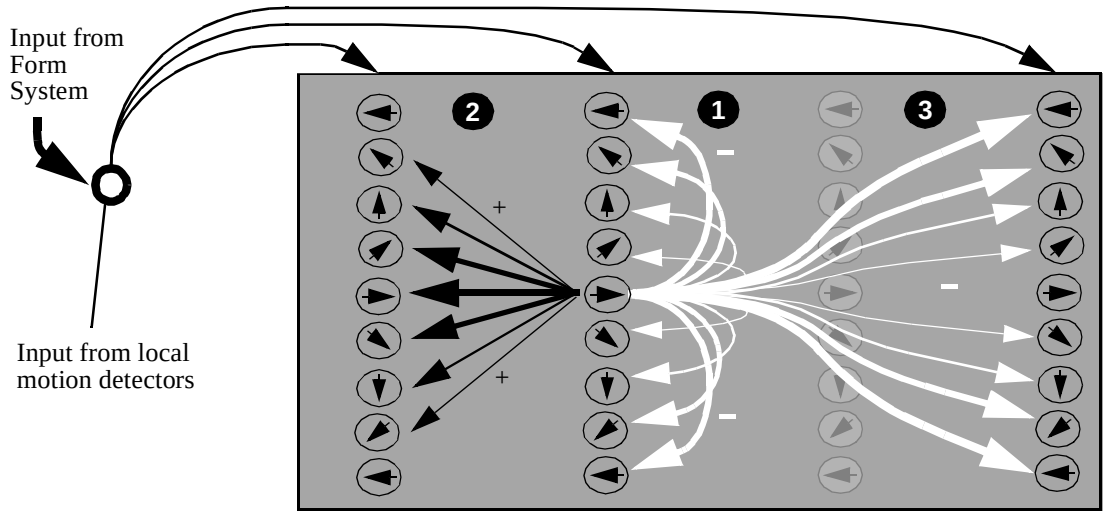

Figure 2 : Integration Cell Architecture: Each position in space in indicated by a vertical column. Note the three processes: [1] inhibition within spatial location across directions, [2] excitation between nearby nodes with similar directional tuning and [3] far-reaching inhibition to nodes with different directional tuning. All connections are weighted by the directional difference.

three sources of excitatory input for the S cells:

[1] *Center-Surround Excitation: S* cells receive center-surround input from the integration cells. They are excited by *I* cells of the same preferred direction in their receptive field center and by *I* cells that possess the opposite preferred direction in their receptive field surround. In this way, *S* cells are excited by direction discontinuities in the pattern of *I* cell activity.

[2] *Excitation from nearby segmentation (S) cells: S* cells also receive excitatory input from other nearby *S* cells of the same directional tuning. This allows for the development of motion borders in regions of the image where there is support for a motion discontinuity.

[3] *Local Motion Detectors (LMDs)*: *S* cells also receive a non-direction specific gating signal from local motion direction cells. A *S* cell cannot be activated by surround input unless it also receives a gating signal from LMD cells or from other *S* cells (See Figure 3). This is in contrast to the *I* system, which allows sub-threshold activity even when there is no underlying activity in the LMDs. The gating mechanism ensures that motion borders only develop in the presence of visual activation or when the existence of such a border is supported by a motion discontinuity in another location where visual

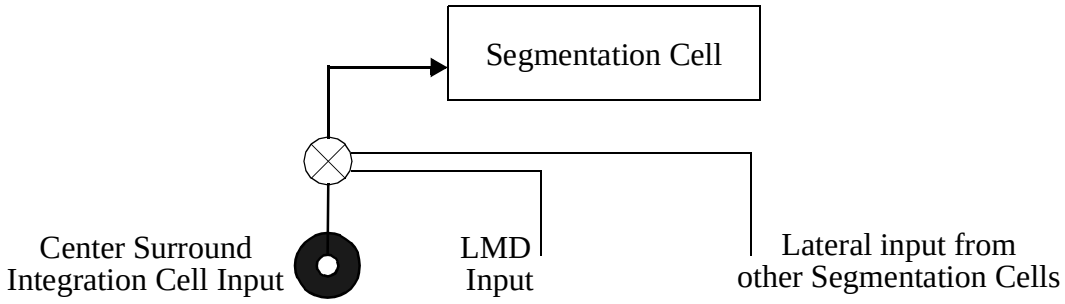

Figure 3: Gating of Segmentation Cell Input: Segmentation cells only receive center-surround input if *both* LMD and lateral segmentation cell inputs are present.

activity exists.

Inhibition: *S* cells are inhibited by *I* cells preferring the same motion direction, located in the surround region of the receptive fields. Unlike the excitatory input, the inhibitory input is not gated. Details of these calculations can be found in Appendix 3.

Segmentation & Integration Systems Interaction

One way in which the *S* and *I* cells interact has already been discussed. Namely, the activity of the *S* cells is determined by *I* cell activity arranged spatially into a center-surround

receptive fields structure. A second type of interaction involves cooperation between *S* cells to form motion borders which serve as barriers to prevent the spread of motion signals by the *I* cells.

The total output of the *S* cells at a given spatial location is used to block the spread of activity of *I* cells at that location. This is achieved through a simple inhibitory connection from each *S* cell to each *I* cell at each spatial location (Appendix 1). Although there is a set of *S* cells representing the full array of preferred directions at each spatial position, the suppression is directionally non-specific, as any discontinuity (regardless of its direction) is relevant. Perceptual analogues to the suppression of motion integration by segmentation processes have been described elsewhere (Watanabe & Cavanagh, 1991). Although such interactions are not strictly necessary to process all motion stimuli, they become critical for processing the motion of multiple overlapping objects in the visual array, since integration should not occur across distinct objects.

Static Form Input

One of the purposes of the current study was to examine how much can be accomplished within the motion system to disambiguate motion signals with the least amount of external input. However, it is not possible to study the motion processing system in isolation, since static form cues clearly play an important role in generating motion percepts. Thus it is also important to consider the influence of interactions between form and motion signals. In the interest of reducing computational complexity, the current model operates only on the *output* of the form system, to examine how such outputs can be used by the motion

system. The network mechanisms by which form information can be generated from retinal input have been modeled elsewhere (Grossberg & Mingolla 1985; Grossberg 1994, Grossberg 1997), and are beyond the scope of this model.

At a minimum, it is critical that any motion processing system suppress motion signals from spatial locations where occlusion is present, as occlusion produces spurious motion signals. When any part of an object passes beneath another, all motion signals orthogonal to the orientation of the occluding contour are lost. Only motion signals parallel to the edge are preserved (Figure 4). Furthermore, such spurious motion signals are unambiguous as the aperture problem only applies when a single edge is present. If such signals were allowed to survive, they might lead to an incorrect disambiguation of object motion.

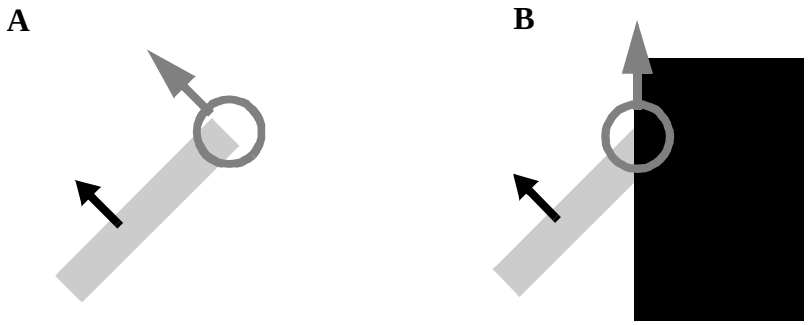

Figure 4: Real terminators unambiguously signal the true direction of motion (A) whereas terminators which result from occlusion unambiguously signal noise (B).

The simulations described herein used T-junctions as indications for occlusion. Rather than simulate the form processing involved in identifying T-junctions, the localization and identification of T-junctions was performed manually. A mask was composed representing the location and position of T-junctions and then used to suppress LMD input to *I* cells in the presence of T-junctions. It is not difficult to imagine a process by which

T-junctions are identified and localized (Guzman 1969).

It is important to point out that T-junctions are not the only cues to occlusion. T-junctions were used in the simulations described herein because they are easy to localize and have a strong impact on motion perception (Cavanagh, 1987; Lidén and Mingolla, 1998). This provided a straightforward means of examining the interaction of form and motion cues. Other cues for occlusion (e.g. accretion/deletion) could also be used to suppress LMD input within the model's motion processing network.

Simulation Techniques - Model Output

The output of the model is conceptualized using a continuous gray-scale semicircle representing motion directions along a 180^{o} continuum. The orientation of the semicircle determines the motion direction indicated by each level of gray. This scheme allows individual regions of the image to be encoded with a particular motion direction, as assigned by the model, and for the evolution of this representation to be displayed in a series of snapshots. This technique was not used for simulations that require the full 360^{o} of motion directions.

Simulation Results

Translating line

As illustrated by the aperture problem (Figure 1), the direction of motion of a translating line is ambiguous along the length of the line. Measurements about the true direction of

motion can only be obtained at the line's terminators. A number of experiments (c.f., Lorenceau, Shiffrar, Wells & Castet 1993) have shown that human perception of translating line segments depends on the duration of stimulus display. For short durations (100 to 400 msec) observers perceive a translating line moving in a direction normal to the line orientation rather than in the veridical direction of motion. Only at longer durations is the line seen to move in the direction of its terminators. Based on these results, it has been proposed that the true direction of line motion is recovered by propagating unambiguous signals from terminators along the contour of the line (Wallach, 1976; Hildreth, 1983; Nakayama & Silverman, 1988a,b; Lorenceau, Shiffrar, Wells, & Castet 1993).

Figure 5 depicts the output of model LMDs for a horizontal line translating at 45° relative to its orientation. Motion of the line is ambiguous along the edge of the line as cells with upward, up-to-the-left and up-to-the-right directional preferences are equally active. Along the terminators, however, unambiguous information about the true direction of motion is available in some locations where only cells coding motion up-to-the-right are active.

The model recovers the true direction of ambiguous line motion by propagating unambiguous motions signals generated by the terminators along the contour of the line. An examination of the evolving activity of the *I* cells demonstrates how the unambiguous motion information available from LMDs centered over the line's terminators is able to disambiguate the motion of the entire line. Initially, the activity of *I* nodes mirrors that of the LMDs (Figure 6**A**). However, over time the activity of nodes at the terminators is enhanced with respect to nodes along the rest of the line (Figure 6B). This is due to the

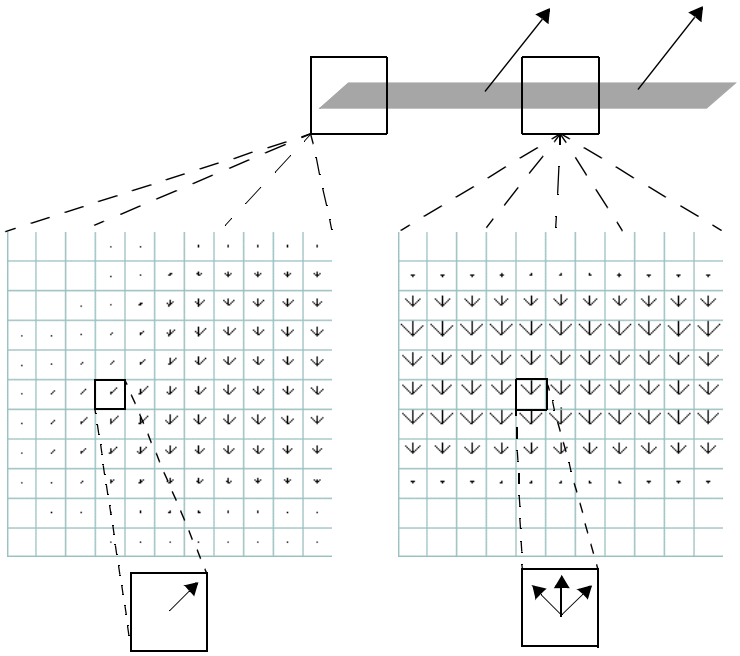

Figure 5: Translating Line - LMD Activation: The grid represents a spatial array of LMD nodes. Each location in the grid contains eight nodes each with a different directional tuning. A 'star' representation is used to depict their activity. The direction of each arm of the star represents a node's preferred direction of motion. The length of the arm represents its activity level. The longer the arm, the greater the activity. Because of the aperture problem, motion is ambiguous anywhere along the edge of the line. Line terminators contain information about the true direction of motion.

directional competition among *I* nodes within each spatial location. The long-range inhibition and short-range excitation interactions of the *I* nodes gradually propagate the unambiguous activation along the length of the entire line. Eventually the activity of all nodes along the line are disambiguated and the entire line is coded as moving up and to the right (Figure 6D).

Line Capture

The introduction of an occluding surface into an image can have a pronounced effect on

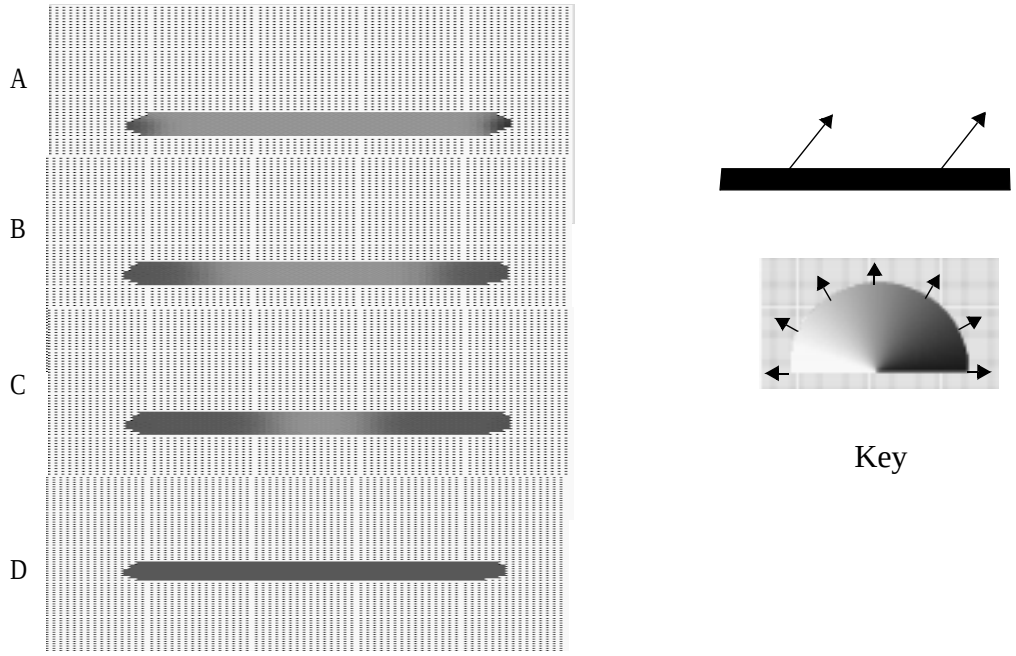

Figure 6: Translating Line - *I* cell activation: The activity of *I* cells in response to a translating line at the different time steps, **A, B, C & D**. Actual motion is up and to the right. The model successfully computes the true motion direction in step D.

the perceived motion within that image. If the terminators of a translating line are obscured, the true motion of the line is undefined. However, the perceived motion of the ambiguous line can be captured by unambiguous motion in a different region of the image, even when the unambiguously moving area is not spatially connected to the ambiguous moving area (Ben-Av & Shiffrar, 1995). Line capture is an example of the more general phenomenon known as motion capture.

Two examples of line capture are shown in Figure 7. The motion of the line on the left (Figure 7A) is ambiguous, since the white diamonds obscure motion of the terminators. However, when the same line is paired with a rightward moving line (Figure 7B) the

perceived direction of the ambiguous line is also rightward. On the other hand, if the ambiguously moving line is paired with an upwardly moving line (Figure 7C), the ambiguous line appears to move upward. The only difference between the stimuli depicted in parts B and C is the motion of the terminators of the paired line; line edge motion is the same in both. Despite the small area of change, the perceived direction of motion of both lines as a whole is noticeably different.

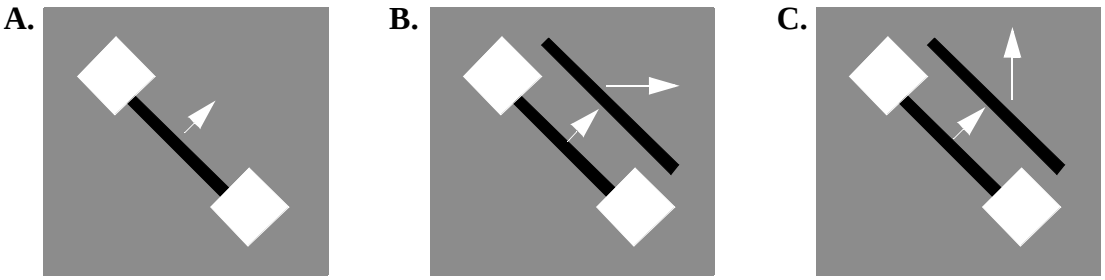

Figure 7 : Line Capture Stimuli. (A) Ambiguous line motion. (B) Horizontal line capture. (C) Vertical line capture. Perceived motion of the ambiguous line is horizontal in (B) and vertical and (C).

Figure 8 shows the evolving activity of *I* cells for the vertical (left column) and horizontal (right column) line capture examples. Initially, as seen in snapshot A, there is little difference in *I* cell activity for horizontal and vertical line capture. The only distinction occurs for the terminators of the non-occluded line, one of which shows rightward motion and the other upward motion. As in the single line example, motion along center of the lines is ambiguous. *I* cell output for the two occluded lines is initially identical. Notice also that there is no motion information at the tips of the occluded lines as LMD input near the occluded terminators is suppressed.

As network activity evolves, disambiguated terminator motion at the ends of the non-

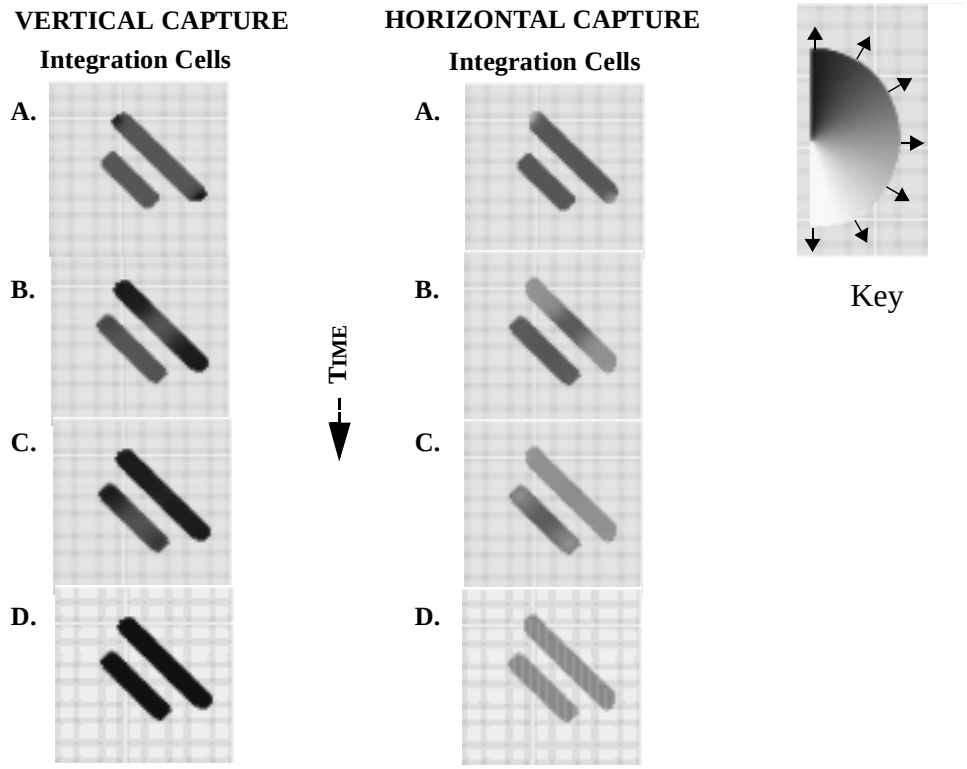

Figure 8 : Line Capture Model Output: Eight snapshots of the evolving sub-threshold and threshold *I* cell activity for vertical (left column) and horizontal (right column) line capture.

occluded bar gradually disambiguates the rest of the non-occluded bar (Figure 8, A-C). The occluded bar remains unchanged, since no disambiguation signal is present (snapshot B). Disambiguating sub-threshold activation from non-occluded bar begins to reach the occluded bar by snapshot C. This signal begins disambiguating the occluded bar. By the final snapshot (D), both bars have been entirely disambiguated. The occluded bars appear to move in the same direction as their non-occluded neighbors.

Divided Diamonds

The previous example showed how unambiguous motion signals can spread to nearby regions in which motion direction is ambiguous. Other experimental evidence suggests

that the spread of unambiguous signals can be blocked by information which reliably signals the existence of a terminator. Lorenceau and Shiffrar (1992) found that subjects exhibited poor performance in judging the direction of rotation of a diamond shaped figure viewed through four (invisible) apertures. However, when terminator information was made ambiguous through changes in contrast, length, or eccentricity, subjects were able to perform the direction discrimination task. Similarly, Shiffrar and Lorenceau (1996) found that decreasing terminator information increased integration across contours.

These effects can be examined using two stimuli. One stimulus is generated by placing opaque bars over a rightward translating diamond (Figure 9A). When the opaque bars are part of the background, the individual parts of the diamond appear to break into four separately moving pieces. The left two pieces appear to move down and up, approaching each other. The right two pieces appear to move up and down, away from each other. The second stimulus employs visible opaque bars (Figure 9D). The perception of motion for the second stimulus is of a rigid structure moving rightward (Figure 9E).

The model's evolving activity supports Lorenceau and Shiffrar's (1992) suggestion that unambiguous motion of terminators inhibits the integration of motion signals across contours. When the occluders are invisible, the model output along the edge of each line is initially in a direction orthogonal to the orientation of the line (Figure 10A - top), as a result of the aperture problem. At the terminators, however, the output follows the direction of motion for the terminators. As in the translating line example, activity of *I* nodes along the edge of the line is suppressed, since motion along the edge is ambiguous and multiple *I* nodes are active. The activity of *I* nodes at the terminators, however, is unam-

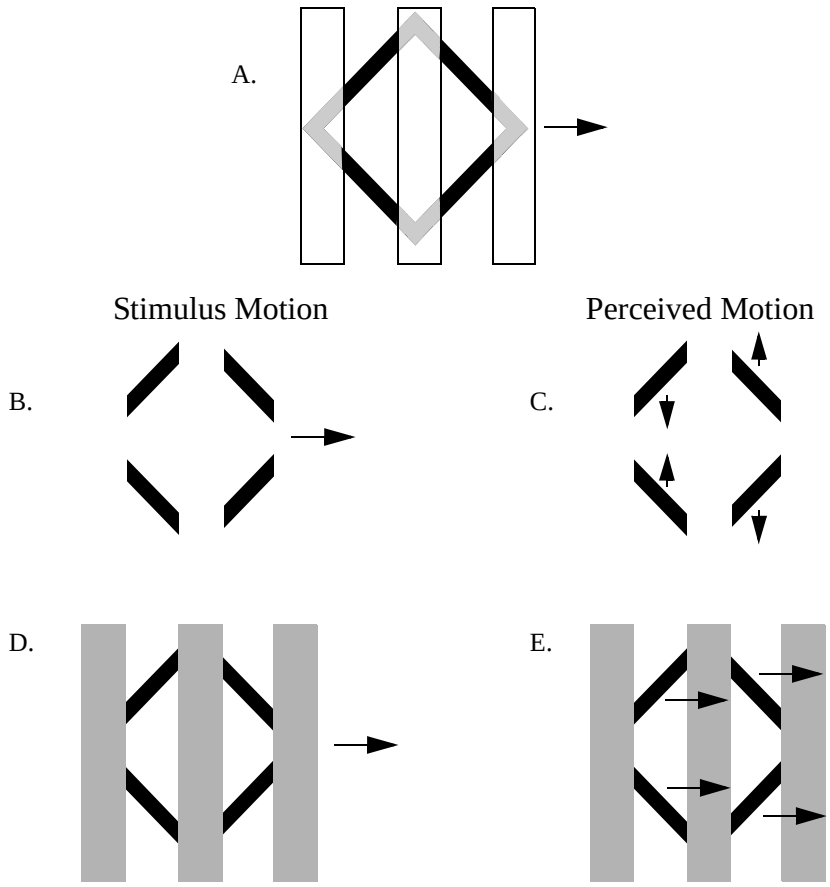

Figure 9: Diamond with Bar Occluders: Stimuli are generated by translating a diamond pattern behind opaque bars. Invisible portions of the diamond are shown in light grey (A). When occluders are invisible, (B), the visible portions of the diamond are seen to move incoherently (C). When visible occluders (D) are employed, the visible portions are seen to move together in the true direction of motion (E)

biguous and thus enhanced. The dynamics of the model lead to a propagation of motion signals, and eventually the output for each line is in the direction of its terminators. Because the terminators for each individual line possess different motion signals, each line is perceived as moving in a different direction (Figure 10C, top).

The results are quite different when the occluders are visible (Figure 9B). Since LMD input to the *I* cells is suppressed when occlusion information (T-junctions) is present,

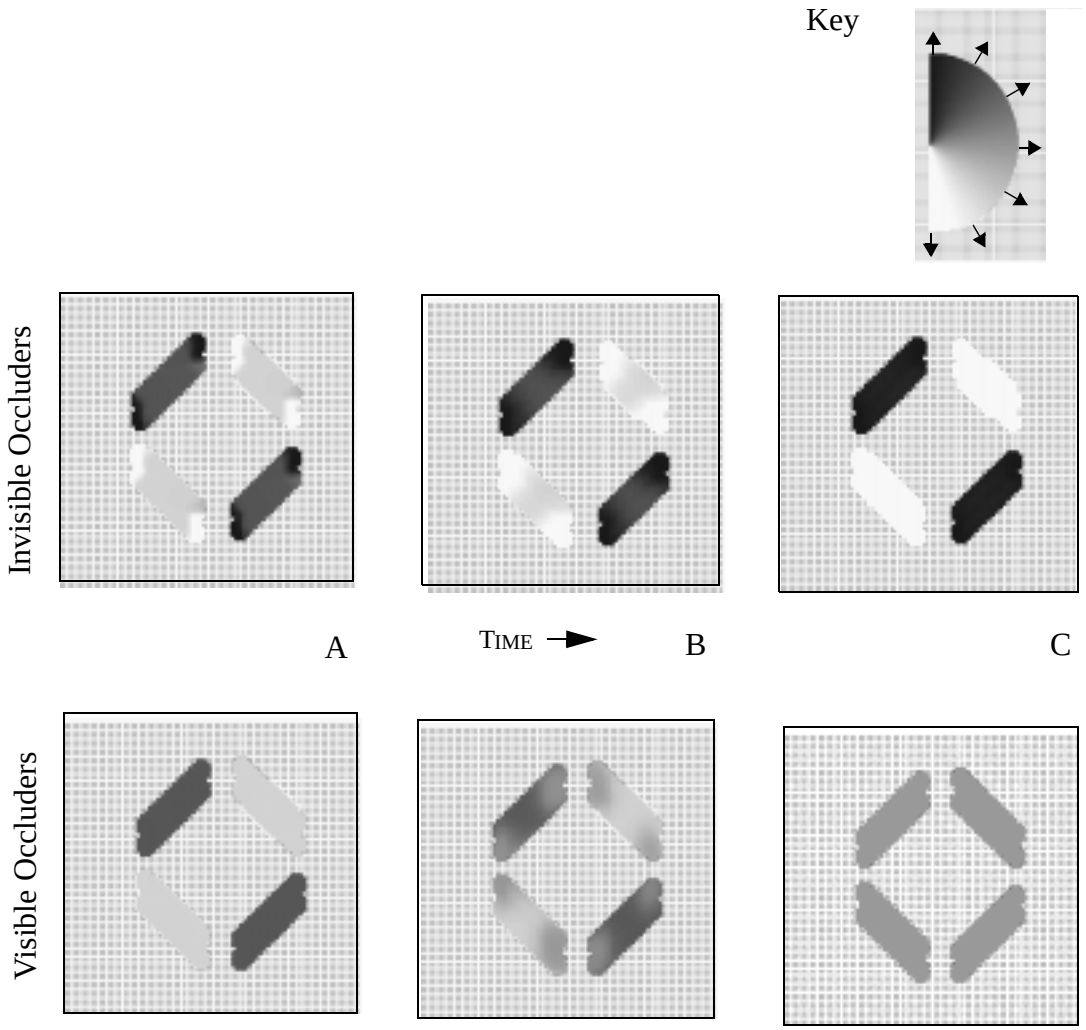

Figure 10: Simulations of divided diamonds. Model output is shown for 3 different time steps. When occluders are invisible (top) the model's output shows the diamond parts moving incoherently. When occluders are visible (bottom) the model signals the true direction of motion. See text for details.

unambiguous motion information at the terminators is no longer present and propagation to the ambiguous centers no longer occurs (Figure 10, bottom). In the simulations discussed thus far, terminators have been the source of unambiguous motion signals. However, unambiguous motion signals can also be generated through the combination of motion signals across space. When motion signals from different areas of space are combined, the competitive dynamics of the model may only allow a subset of the combined

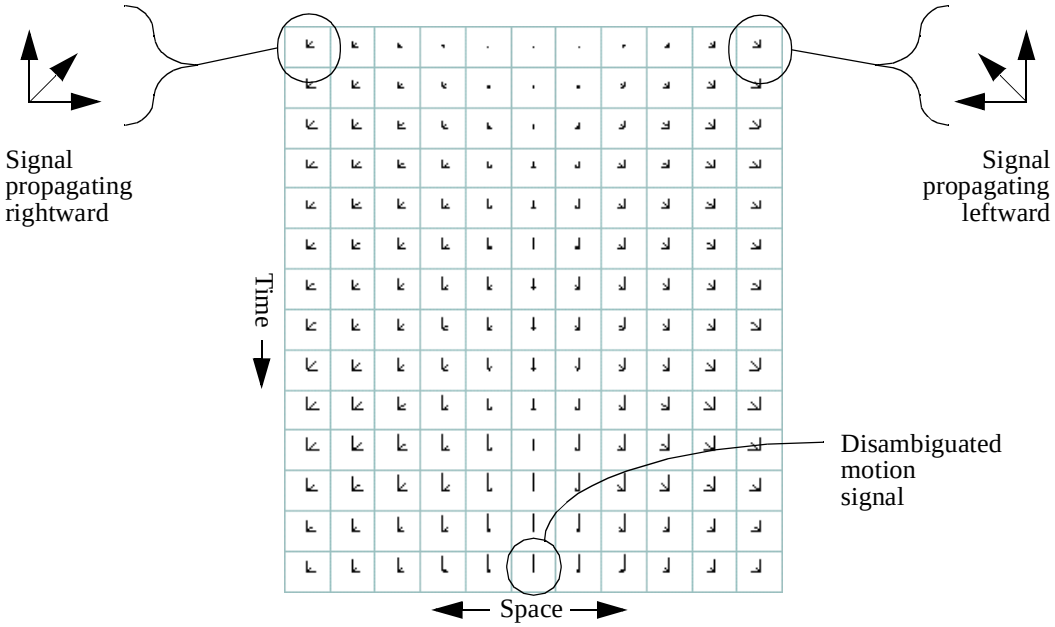

Figure 11: When two ambiguous motion signals from different areas of space propagate towards each other, a smaller subset of motions may be compatible with both. In the pictured example, two motion signal meet (one coming from the left and one from the right) and only upward motion is compatible with both.

motions to survive (Figure 11). In the diamond example, only rightward motion is compatible with the motion signals propagating from all four lines. This new rightward unambiguous signal serves to disambiguates the motion of all four line segments. The model's behavior is in agreement with Lorenceau and Shiffrar's (1992) suggestion that the integration of motion across contours is facilitated when terminators are ambiguous.

Crossing Lines

The motion simulations presented thus far have involved only single objects or cases in which there were no discontinuities in the motion signals arising within objects. Although segmentation (*S*) cells were present in the model, they played an insignificant role as their

activation was minimal. This section presents an example in which the activity of the segmentation cells is critical in explaining the perceptual effect.

One of the most fundamental and difficult problems in motion processing occurs when multiple conflicting motion signals are present in an image. A simple example of this problem is what is known as 'the cross problem' (Weiss & Adelson, 1994), which arises when two orthogonal lines translate horizontally past each another (Figure 12A). This is also referred to as the 'chopstick illusion.' (Anstis, 1990). Unlike in the previous examples, there are two conflicting sources of unambiguous motion information in the display. Motion signals at the line terminators indicate the true direction of motion, while motion at the intersection signals upward motion.

A variation on the cross problem occurs when occluding patches are added to the top and bottom of the moving lines (Figure 12C). When occluders are present, the veridical motion of the lines is no longer perceived. Instead, the two lines appear to move upward as a single rigid object (Figure 12D). Taken together, these phenomena suggest that the visual system relies on terminator information when it is reliable, and intersection information when it is not.

The model segmentation system uses cells with center-surround receptive fields to discover motion discontinuities in the image. Their activity in turn suppresses the activity of the *I* cells at nearby locations, preventing the integration of motion signals in the presence of motion discontinuities. For the cross stimulus, there are strong motion discontinuities at the intersection of the two lines (Figure 13). *S* cells are activated by this discontinuity and subsequently suppress *I* cell activity near the intersection. This suppres-

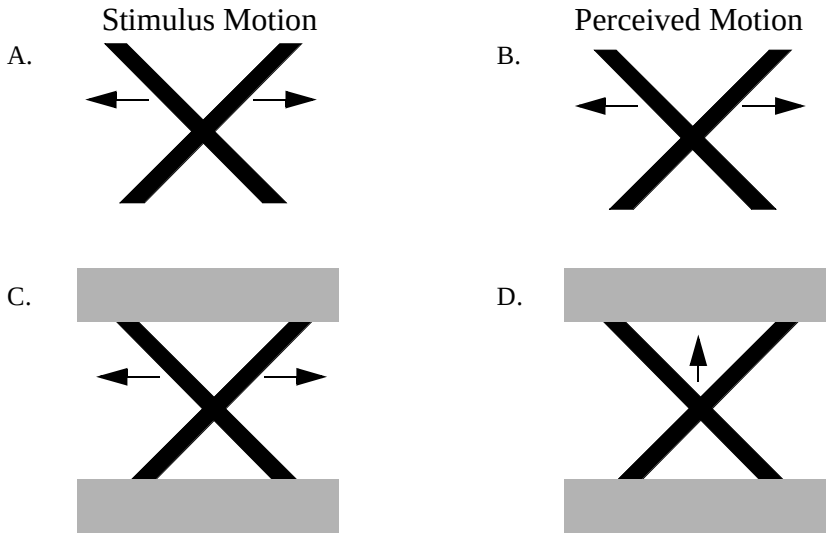

Figure 12 : Crossing Lines: Stimuli consist of two intersecting lines translating in opposite directions When occluders are absent (A), the visible portions of the diamond are seen to move in their veridical direction (B). When occluders are present (C), the lines are seen to move coherently upward as a single rigid object (D).

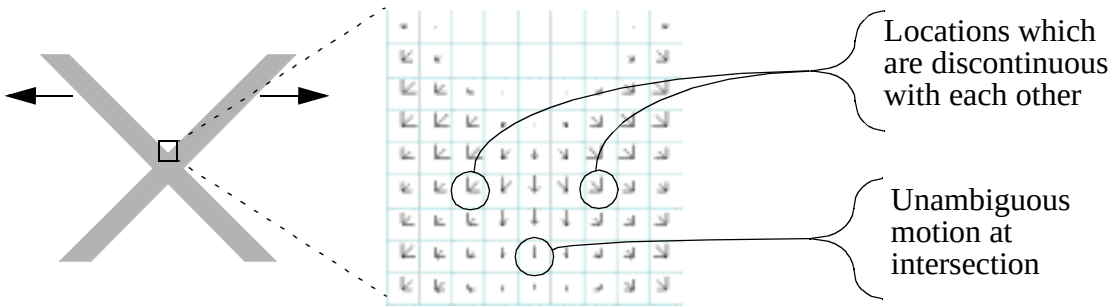

Figure 13 : Initial motion measurements from LMDs are shown for a small portion of the cross stimulus. Unambiguous upward motion signals at the intersection of the cross stimulus are accompanied by nearby motion discontinuities.

sion prevents unambiguous upward motion signals at the intersection from propagating outward. *S* cells which receive propagated input signals from sub-threshold *I* cells are also activated via the gating signal from active *S* cells. This allows a motion border to grow outward separating the different motion areas and preventing the integration of sub-threshold *I* cells. Consequently, unambiguous motion signals propagating inward from the line

terminators eventually dominate the network activity. The evolving activity of the *S* and *I* cells for the unoccluded cross stimulus is depicted in the left two columns of Figure 14. *S* cell activity is depicted in the first column. *I* cell activity, as shown in the second column, is suppressed at locations where *S* cells are active. The model output depicts the two lines moving in separate directions.

When occluders are present (Figure 14, right two columns), unambiguous motion signals from the terminators are unavailable. As the intersection is disambiguated, the motion discontinuity disappears and consequently *S* cell activity is reduced. This allows *I* cells at the intersection to be active. The final model output depicts a rigid cross moving upwards.

Discontinuities within an object

The previous example demonstrated how the localization of motion discontinuities prevents spurious integration between different objects. In some cases, however, discontinuities are present within a single rigid object. For example, a translating diamond or rectangle contains motion discontinuities near each corner (Figure 15) However, the perceived object motion, when properly disambiguated should consist of one, not four separate pieces.

The model readily handles such cases. For the example given in Figure 15, initially motion along each side of the square is locally ambiguous. Discontinuities between edges near the corners activate the *S* cells. Segmentation borders grow inward along each edge as sub-threshold activation is propagated over space. However, unambiguous motion sig-

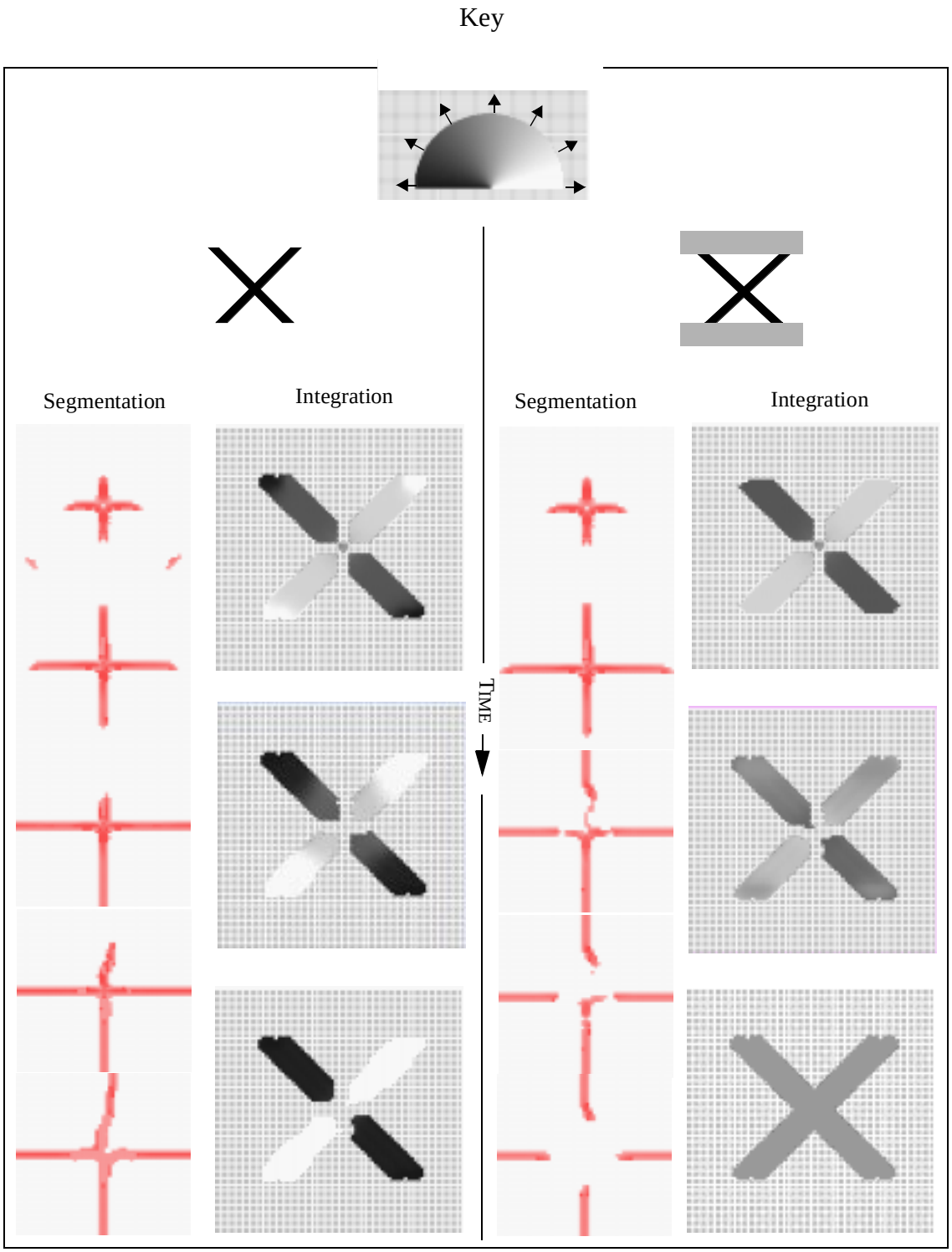

Figure 14 : Model output for the non-occluded (left) and occluded (right) cross stimulus. Both segmentation and integration node activity is shown. See text for details.

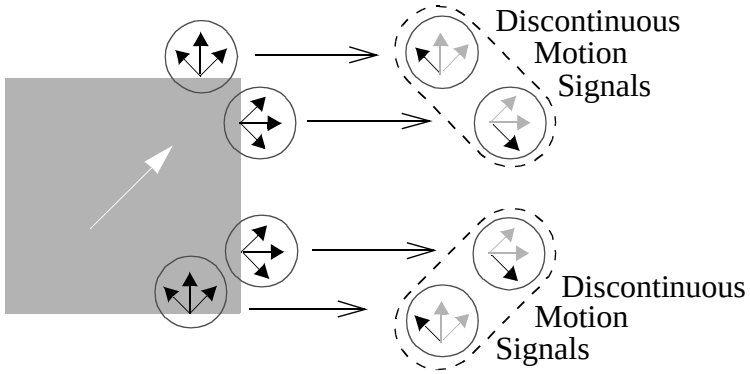

Figure 15: Within-Object Discontinuities: Many objects, such as a translating rectangle, contain discontinuities. In the pictured example, motion discontinuities exist at each of the four corners of the rectangle (discontinuities are illustrated at only two of the corners).

nals at each corner cause nearby *S* cell activity to collapse. The sides are disambiguated and finally, motion borders collapse entirely, and the entire square is disambiguated as one moving object.

The Diamond Problem: What to Disambiguate?

An especially difficult problem for motion processing models occurs when multiple objects move in different directions in contiguous locations in space. For instance, when two diamond-shaped objects pass each other in the visual field, each diamond contains motion which is locally ambiguous, as a consequence of the aperture problem. For the case of a single object, it is sufficient to integrate motion signals near the edges to obtain the true global motion direction. However, when two objects are present, it is important to avoid integrating across objects (Figure 16). For example, when an upward and down-

ward moving diamond pass each other there are various ways in which the motion signals can be integrated. If the left and right motion pairs are combined, a correct disambiguation results. On the other hand combining the top and bottom motion pairs results in spurious motion directions.

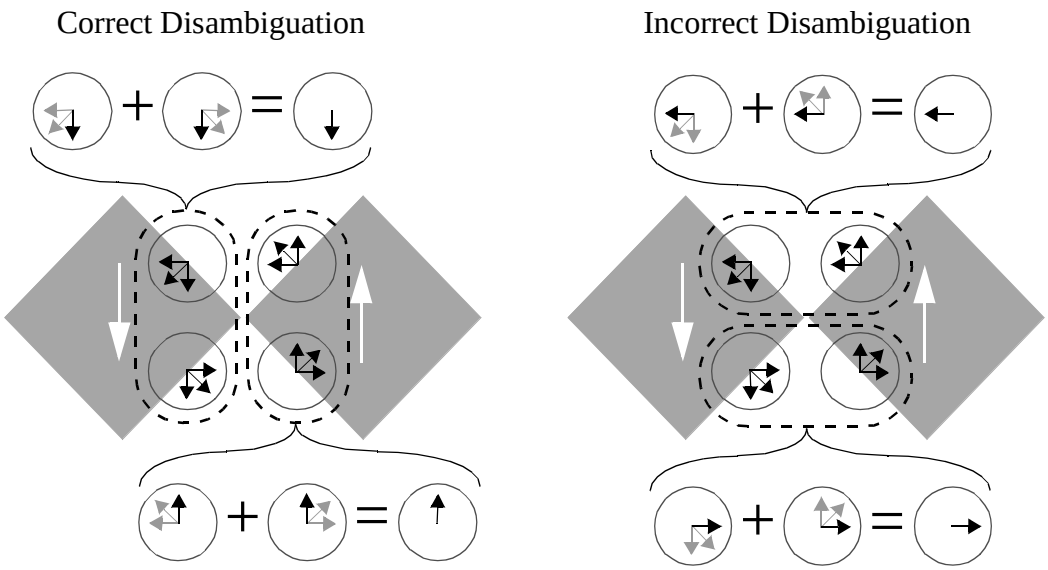

Figure 16: Passing Diamonds Problem: The motion of a diamond can be disambiguated by combining ambiguous edge information. However without knowledge of the underlying shape structure, there is no obvious way to know which signals should be combined. The combination on the left reveals the true motion of the diamonds (up and down). The combination on the right results in spurious disambiguation (left and right).

The model demonstrates that the passing diamond problem is overcome when terminator motion and motion discontinuities are taken into account (Figure 17). As in the previous example, *S* cells are initially activated by motion discontinuities at the corners of each object (Figure 17A). The discontinuity between the diamonds begins to activate other *S* cells between the two diamonds (B-C). Disambiguation results in a collapse of *S* cell activity as motion discontinuities in the corners fade (D). At the same time, segmen-

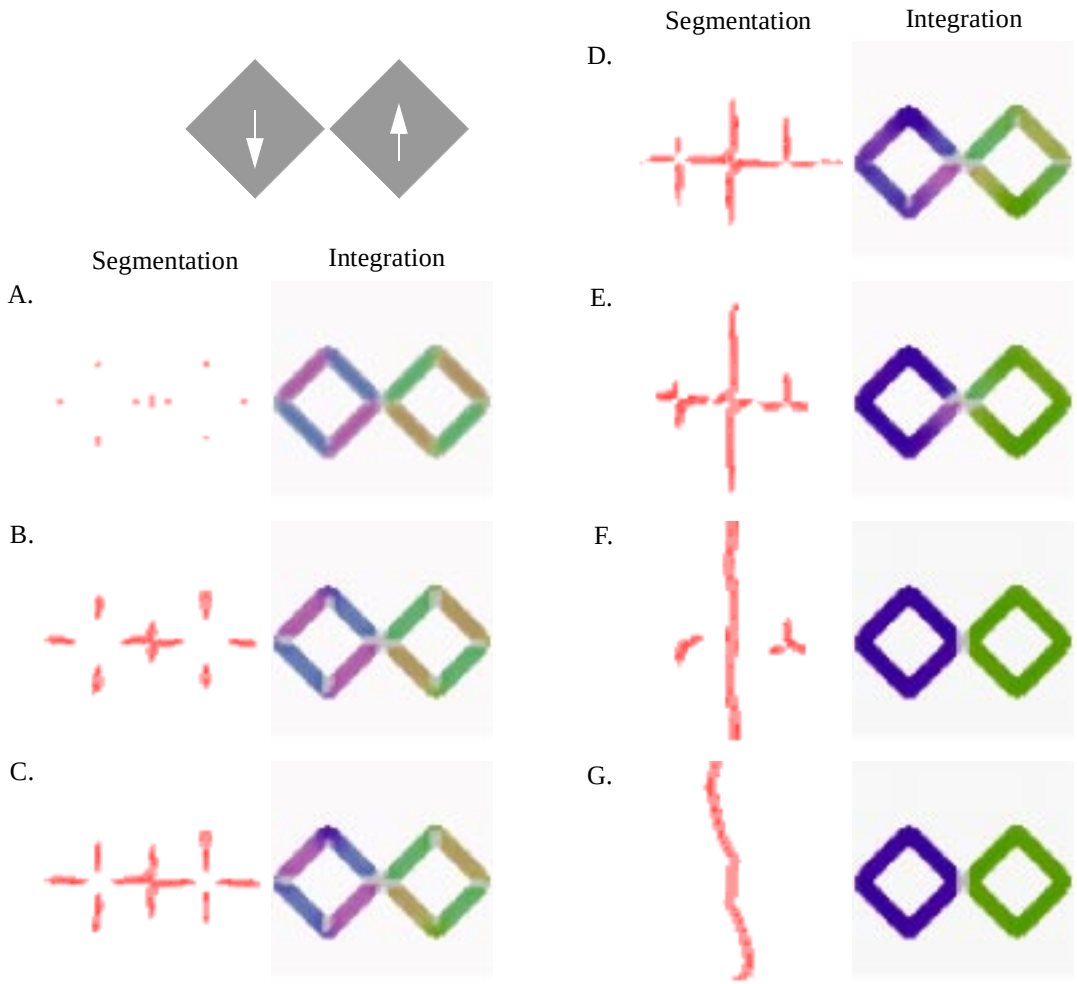

Figure 17 : Two Passing Diamonds: Evolving network activity for the segmentation and integration cells is shown for two passing diamonds. The increasing activity in the segmentation system prevents integration of motion signals between objects. Note that unlike previous diagrams figure lightness does not represent direction.

tation activity along the border between the two diamonds continues to grow as no disambiguating signal is present (E). Finally both diamonds are disambiguated. The remaining active segmentation cells divide the space into two separate regions, one for each diamond (G).

Multiple Object Displays

This section demonstrates how the model can cope with multiple overlapping objects, using an example consisting of two approaching rectangles (Figure 18A). The initial motion measurement by the LMDs at each line-edge pair is distinct and ambiguous. Furthermore, no motion signals exist in the interior of the two objects as their surfaces are uniform (Figure 18B). Correctly disambiguating the motion of these two objects is critically dependent on two mechanisms (Figure 18C). First, spurious motion signals from LMDs near the T-junctions at the intersection of the two objects must be suppressed. This was accomplished as before by constructing a mask at the location of T-junctions in the image. Secondly, the segmentation system must suppress *I* cell activity in the presence of motion discontinuities.

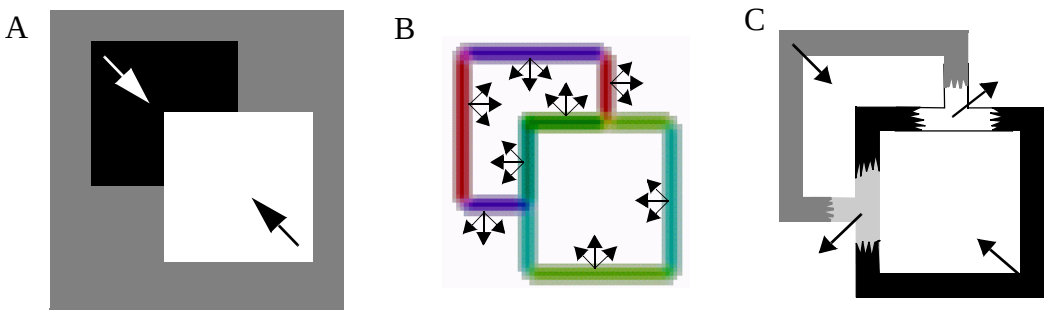

Figure 18: Two approaching rectangles. For this input (A), LMD output (B) contains spurious motion directions. When either T-junction masking (for occlusion) or the segmentation system is not employed, the model's output for multiple objects is incorrect (C).

Figure 19 shows how the model is able to segment the motion of the two rectangles. Initially (snapshots A) *S* cells are activated by discontinuities at the corners of the rectangles and at the intersections between the rectangles. As sub-threshold motion signals are propagated across the image, motion borders grow inward to the center of and between the

two rectangles (snapshots B). When the corners of the rectangles are disambiguated, the motion discontinuity disappears, and *S* cells at the corners are deactivated. Motion borders within the rectangles are no longer supported and begin to degrade (snapshots C). By the final snapshot, both rectangles have been completely disambiguated. Note that the remaining segmentation border divides the space into two regions containing the two different object motions (D).

Summary

A motion processing model with sub-systems for integration and segmentation can interpret a wide range of motion stimuli. The integration system serves to overcome the aperture problem and to eliminate noise in the image. The segmentation system prevents the integration of motion signals between objects. Despite their seemingly contradictory roles, the model demonstrates how they complement each other to produce a unified interpretation of motion stimuli. While the integration system is sufficient to explain human motion perception for the barber pole illusion (Lidén, 1998), motion capture and crossing lines, both the segmentation and integration systems are required for stimuli involving multiple objects with different trajectories.

Discussion

The model presented in this paper consists of two computational processes implemented by two separate, but interacting sets of nodes. It explains how two seemingly contradictory processes, integration and segmentation, can interact to create a unified interpretation

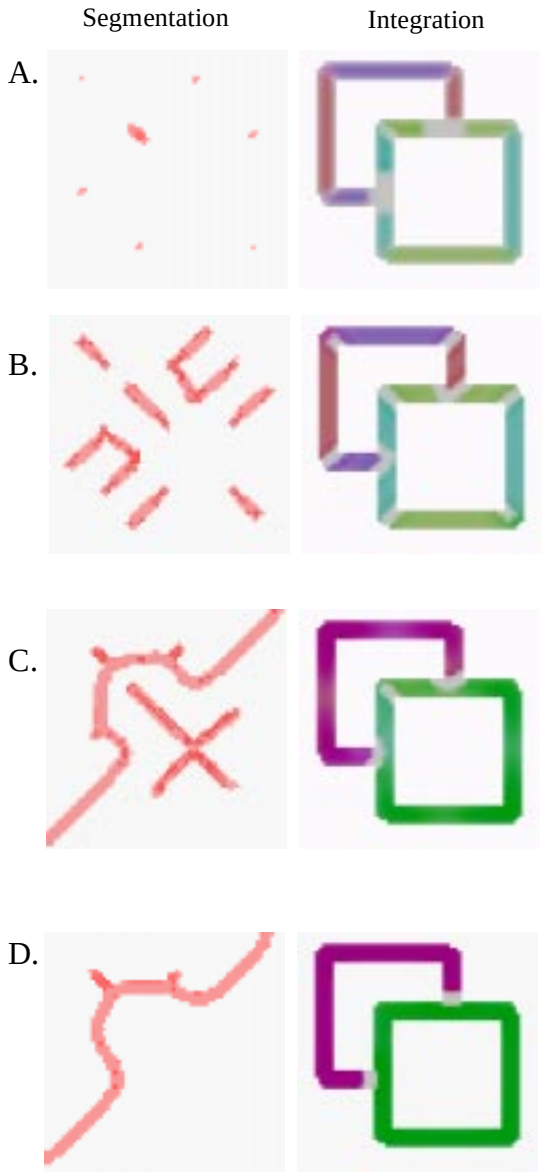

Figure 19: Four snapshots of segmentation and integration cell output for two moving rectangles. Segmentation cell activity blocks the integration of motion signals across objects. See text for details.

of visual motion. Integration nodes smooth over noise in the input image, overcome the aperture problem, and propagate motion signals across visual space. Segmentation nodes detect motion discontinuities through the use of motion center-surround receptive fields. The interaction of these two processes with local motion signals explains how global motion percepts evolve over time. Simulations demonstrate that the model explains a

number of perceptual illusions as consequences of mechanisms which compensate for noise and ambiguity in the measurement of local motion.

Neurophysiological Interpretation

Local Motion Detection

The initial extraction of motion signals takes place in the primary visual cortex (V1). Layer 4b of the primary visual cortex contains a particularly large number of direction-selective neurons (Dow, 1974; Blasdel & Fitzpatrick 1984). These cells could fill the roles of the local motion detectors (LMDs) in the model. Direction-selective cells in V1 give rise to the greatest number of projections to the middle temporal (MT) area (Maunsell & Van Essen 1983). MT also receives input from thick stripes in the secondary visual cortex (V2) where a large proportion of neurons are direction-selective (DeYoe & Van Essen, 1985; Shipp & Zeki 1985).

Integration and Segmentation in MT

Rodman & Albright (1989) suggest that the middle temporal cortical area (MT) is the first cortical area in which the individual components of a complex stimulus are combined to form a representation of the global motion. Similarly, Qian & Anderson (1994) argue that MT is the first cortical region in which the suppression of local motion signals is employed to achieve a reduction in the noise of a motion stimulus. Area MT would therefore seem to be the first likely candidate for a neural area capable of simultaneously performing integration and segmentation of motion signals. MT is comprised of motion

selective cells (Zeki, 1974a&b) with a full topographical representation of the visual field (Gattass & Gross, 1981). Inhibition between cells encoding opposite directions, as hypothesized in the current model, are important for generating MT response properties (Snowden, Treue, Erickson & Andersen, 1991; Bradley et al., 1995).

Physiological studies of MT reveal two cell types analogous to the integration and segmentation cells used in the model (Allman, Miezin & McGuinness, 1985; Tanaka, Hikosaka, Saito, Yukie, Fukada & Iwai, 1986; Saito, 1993). Cells in one group respond to a bar moving in a specific direction in the receptive field center, and are suppressed by a dot pattern moving in the same direction in the surround. These cells are similar, in terms of response properties and inhibitory receptive field surrounds, to the segmentation *(S*) cells in the model. Another class of cells has no surround suppression, and show increasing responses to increasing stimulus sizes. These cells are similar to the integration (*I*) cells in the model. Born and Tootell (1992) found that these two cell types are segregated in a columnar fashion in MT.

Anatomical studies of MT report numerous fibers oriented laterally with respect to cortical surface (Van Essen, Maunsell, Bixby 1981). These connections could serve as a basis for the recurrent spatial interactions used in the model. Other anatomical work has demonstrated that inhibitory surrounds in MT segmentation cells are likely to arise from connections among MT cells, rather than from the pattern of inputs from V1, since cells with center-surround receptive field structure in the input layer of MT are rare (Raiguel, Van Hulle, Xiao, Marcar & Orban 1995). This provides further support for the model's use of recurrent connectivity in shaping receptive field structures.

It is also known that opponent cells in MT are capable of signalling a kinetic boundary, as would be important for the segmentation nodes in the model (Marcar, Xiao, Raiguel, Maes & Orban, 1995). However, these MT cells cannot accurately encode the orientation of a kinetic boundary or the position of the boundary within their receptive fields. The model, however, demonstrates that such coding of orientation and position within the receptive field of segmentation cells is not necessary for object segmentation. It is only necessary for the *presence* of a motion discontinuity to suppress the activity of integration nodes. Such suppression in the model is directionally non-specific and spatially diffuse.

Global Motion Processing in MST

One of the major output areas for MT is another cortical area known as the medial superior temporal (MST) area (Maunsell & Van Essen, 1983). MST appears to be further subdivided into at least two areas (Komatsu & Wurtz, 1988). The ventral part of MST (MSTv) contains cells which respond best to small fields of motion in a preferred direction, and are inhibited by large-field motion (Tanaka, Sugita, Moriya & Saito, 1993). Cells in the dorsal part of MST (MSTd) respond to large motion stimuli, and are presumably used for the analysis of self-motion (Wurtz, Yamasaki, Duffy & Roy, 1990; Duffy & Wurtz, 1991a&b) or complex object motion (Graziano, Andersen & Snowden, 1994; Geesaman & Anderson, 1996; Ferrera & Lisberger, 1997), or both (Pack, 1998). Thus, it appears that the distinction between motion integration and segmentation is preserved at the next stage in the cortical motion processing hierarchy. Theoretical work has shown

that this distinction is important for controlling smooth pursuit eye movements (Pack et al., 1998).

Comparison with Other Models

Segmentation or *Integration*

The processes of integration and segmentation have contradictory goals. Integration eliminates local motion signals creating uniform motion in a region. Segmentation, on the other hand, enhances local motion differences within a region. While previous models have focused on one of these computations, the current model demonstrates how the two processes can be combined.

Integration Models

Models that perform motion integration often fail to achieve adequate image segmentation. One method for overcoming noise and the aperture problem is to integrate motion signals over a spatial neighborhood of a fixed distance (Hildreth and Koch, 1987; Bulthoff, Little & Poggio, 1989a & 1989b), but this approach does not account for distinct objects moving in nearby regions of space. As a result incorrect motion estimates often occur when integration proceeds across objects. Regularization methods (Horn & Schunck, 1981; Yuille & Grzywacz, 1988; Wang, Mathur, &Koch, 1989; Koch, Wang & Mathur, 1989) are particularly susceptible to smoothing over multi-image displays (Poggio, Torre & Koch. 1985).

Some modified regularization models come closer to addressing both segmentation

and integration. Hutchinson, Koch, Luo and Mead (1988) presented a hardware implementation of a regularization approach which specifically allows for the smoothness assumption to be discarded across adjacent regions when there is evidence for the presence of a motion discontinuity. This is accomplished by incorporating information about luminance edges from zero-crossings and motion discontinuities. However, the relationship between modified regularization processes and human motion perception is unclear (Braddick, 1993)

Segmentation Models

Nakayama & Loomis (1974) propose the idea of a *convexity function* which compares motion direction over its center and surround in different directions. Other models, such as that of Sachtler & Zaidi (1995) use similar center-surround receptive fields to detect motion discontinuities. These center-surround models are limited in scope to motion segmentation. As a result, they do not address important problems which are solved by integration approaches, including the aperture problem.

Another way to view the problem of resolving motion ambiguities and the detection of motion boundaries involves the propagation of constraints on motion signals. Marr (1982) noted that although the aperture problem results in ambiguous motion signals, it also provides a range of motions which are inconsistent with the local motion. By combining these so-called “forbidden” directions of motion with forbidden directions in neighboring locations, a smaller subset of motions can be found which are compatible with the two areas. If combination with a neighboring region results in all motions being forbidden,

it is likely that the initial measurements of the two regions belong to different objects. Although such a method might be effective for performing motion segmentation in a limited domain, like center-surround models, it fails to address motion integration.

One model which achieves a crude segmentation is that of Nowlan & Sejnowski (1994). Their filter selection model segments an image into multiple regions of support for each motion in the image. However, it does so by ignoring bad estimates of motion. Thus, the support regions for each object are a small fraction of the local-velocity measurements. The final output of the model has no spatial resolution as it consists of one set of units which represents the entire visual field. Some additional process would be required for true object segmentation.

Chey, Grossberg & Mingolla (1997) present a model that extracts initial motion measurements, accounts for much of the psychophysical data on speed tuning, and some data on the coherence and incoherence of plaid patterns. However, the model's segmentation system works only with static form input, and does not contain a mechanism for segmenting motion discontinuities. As a result, it does not compute the motion of multiple objects in a scene.

Temporal Aspects of Integration and Segmentation

A second criticism of most motion models is that they fail to deal with the temporal aspects of motion integration and propagation. The importance of temporal integration is lost on the regularization models discussed in the previous section. Such methods fail to explain how humans can track the trajectory of a single dot among a field of moving ran-

dom dots (Watamaniuk, McKee & Grzywacz 1995). Regularization results in the obliteration of a single signal dot in noise.

Of the models discussed thus far, only the model of Chey et al. (1997) includes the potential to model some important aspects of the temporal integration of motion. The model incorporates neurons which act as leaky integrators, summing their inputs over time, as well as lateral recurrent feedback. Together these mechanisms allow motion propagation to occur, notably from feature tracking locations that compute unambiguous motion directions to locations where only ambiguous computations occur due to the aperture problem.

Future Work

A number of important issues in motion processing have not been addressed by the current model. First, the masking of local motion cues in the presence of T-junctions could be expanded to include other cues to occlusion. Furthermore, rather than using an all-or-none masking process, the strength of local motion suppression could reflect the strength of the occlusion cue. This would be useful for studying phenomena such as the barber pole illusion, in which local T-junctions caused by the intersection of individual grating bars with an occluding surface have a minimal impact relative to the T-junctions caused by the intersection of the entire grating patch with the occluder (Shimojo et al., 1989; Lidén & Mingolla, 1998). This issue has been addressed in other work (Lidén, 1998).

In addition to utilizing occlusion information, it would be useful to model the function of binocular disparity in motion processing. Bradley and Andersen (1995) found that,

for many MT cells, the inhibition between cells encoding opposite directions is strongly modulated by disparity. Specifically, the inhibition is stronger between cells that encode similar disparities. This suggests a natural extension of the model to three dimensions, wherein binocular disparity inputs modulate the inhibition between cells encoding opposite directions. Such a scheme could be used to preserve the smoothing aspects of the integrative system, while segmenting the motion of different objects or surfaces according to their depths. An interesting test of this approach would be to examine if depth could be perceived from transparent motion stimuli, as occurs even when no disparity cues are present (Hiris and Blake, 1996).

Conclusion

The model presented in this paper demonstrates how a number of puzzling perceptual phenomena emerge as a consequence of interactions among neural units attempting to create a globally coherent motion percept from a possible noisy image representation. These interactions consist of processes which integrate and segment motion information in the image, and account for depth interactions. In particular, occlusion cues such as T-junctions are shown to be crucial to the understanding of motion perception. Simulations indicate that model cell activity evolves over time and space in a manner that is consistent with human motion perception. It therefore seems likely that the model can be used to guide future experimental work.

References

Adelson, E. & Bergen, J. (1985). Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America*: A, 2(2), 284-299.

Adelson, E. & Movshon, A. (1982). Phenomenal coherence of moving visual patterns. *Nature*, 300, 523-525.

Allman, J., Miezin, F. & McGuinness, E. (1985). Direction- and velocity-specific responses from beyond the classical receptive field in the middle temporal visual area (MT). *Perception*, 14, 105-126.

Anstis, S. (1990) Imperceptible intersections: The chopstick illusion. In Blake, A. & Troscianko, T. (Ed.), *AI and the Eye*, John Wiley & Sons Ltd., New York, pp.105-117.

Barlow, H. & Levick, W. (1965). The mechanism of directionally selective units in rabbit's retina. *Journal of Physiology*, 178, 477-504.

Ben-Av, MB & Shiffrar, M. (1995) Disambiguating velocity estimates across image space. *Vision Research*, 35, 2889-2995.

Blasdel, G. & Fitzpatrick, D. (1984). Physiological organization of layer 4 in macaque striate cortex. *Journal of Neuroscience*, 4(3), 880-895.

Born, R. & Tootell, R. (1992). Segregation of global and local motion processing in primate middle temporal visual area. *Nature*, 357, 497-499.

Braddick, O. (1993). Segmentation versus integration in visual motion processing. *Trends in Neurosciences*, 16(7), 263-268.

Bradley, D., Qian, N. & Anderson, R. (1995). Integration of motion and stereopsis in middle temporal cortical area of macaque. *Nature*, 373, 609-611.

Bulthoff, H., Little, J. & Poggio, T. (1989a). A parallel algorithm for real-time computation of optical flow. *Nature*, 337, 549-553.

Bulthoff, H., Little, J. & Poggio, T. (1989b). A parallel motion algorithm consistent with psychophysics and physiology. *Proceedings: Workshop on Visual Motion*, Irvine, California, 165-172.

Castet, E., Charton, V., & Dufour A. (In Press). The extrinsic/intrinsic classification of 2D motion signals with barber-pole stimuli. *Vision Research.*

Cavanagh, P. (1987). Reconstructing the third dimension: Interactions between color, texture, motion, binocular disparity, and shape. *Computer Vision, Graphics, and Image Processing*, 37, 171-195.

Chey, J., Grossberg, S. & Mingolla, E. (1997). Neural dynamics of motion grouping: From aperture ambiguity to object speed and direction. *Journal of the Optical Society of America*, 14(10), 2570-2594.

DeYoe, E. & Van Essen, D. (1985). Segregation of efferent connections and receptive field properties in visual area V2 of the macaque. *Nature*, 317, 58-61.

Dow, B. (1974). Functional classes of cells and their laminar distribution in monkey visual cortex. *Journal of Neurophysiology*, 37, 927-945.

Duffy, C. & Wurtz, R. (1991a). Sensitivity of MST neurons to optic flow stimuli. I. A continuum of response selectivity to large-field stimuli. *Journal of Neurophysiology*, 65(6), 1329-1345.

Duffy, C. & Wurtz, R. (1991b). Sensitivity of MST neurons to optic flow stimuli. II.Mechanisms of response selectivity revealed by small-field stimuli. *Journal of Neurophysiology*, 65(6), 1346-1359.

Fennema, C. & Thompson, W. (1979). Velocity determination in scenes containing several moving objects. *Computer Graphics and Image Processing*, 9, 301-315.

Ferrera, V. & Lisberger, S. (1997). Neuronal responses in visual areas MT and MST during smooth pursuit target selection. *Journal of Neurophysiology*, 78, 1433-1446.

Gattass, R. & Gross, C. (1981). Visual topography of striate projection zone (MT) in posterior superior temporal sulcus of the macaque. *Journal of Neurophysiology*, 46(3), 621-638.

Geesaman, B. & Andersen, R. (1996). The analysis of complex motion patterns by form/cue invariant MSTd neurons. *Journal of Neuroscience*, 16(15), 4716-4732.

Graziano, M., Andersen, R. & Snowden, R. (1994). Tuning of MST neurons to spiral motions. *Journal of Neuroscience*, 14(1), 54-67.

Grossberg, S. (1994). 3-D vision and figure-ground separation by visual cortex. *Perception & Psychophysics*, 55(1), 48-120.

Grossberg, S. (1997) Cortical dynamics of three-dimensional figure-ground perception of two-dimensional pictures. *Psychological Review*, 104(3), 618-658.

Grossberg, S. & Mingolla, E. (1985). Neural dynamics of form perception: Boundary completion illusory figures and neon color spreading. *Psychological Review*, 92(2), 173-311.

Grossberg, S. & Rudd, M. (1989). A neural architecture for visual motion perception: Group and element apparent motion. *Neural Networks*, 2, 421-450.

Grossberg, S. & Rudd, M. E. (1992). Cortical dynamics of visual motion perception: Short-range and long-range apparent motion. *Psychological Review*, 99(1), 78-121.

Grunewald, A. (1996). A model of transparent motion and non-transparent motion aftereffects. *Advances in Neural Information Processing Systems*, 838-843.

Guzman, A. (1969). Decomposition of a visual scene into three-dimensional bodies. In Grasselli, A. (Ed.), *Automatic Interpretation and Classification of Images* (pp. 243-276). New York: Academic Press.

Hildreth, E. (1983). The measurement of visual motion. MIT Press, Cambridge, MA.

Hildreth, E. & Koch, C. (1987). The analysis of visual motion: From computational theory to neuronal mechanisms. *Annual Reviews in Neuroscience*, 10, 477-633.

Hiris, E. & Blake, R. (1996) Direction repulsion in motion transparency. *Visual Neuroscience*, 13, 187-197.

Horn, B. & Schunck, B. (1981). Determining optical flow. *Artificial Intelligence*, 17, 185-203.

Hutchinson, J., Koch, C., Luo, J. & Mead, C. (1988). Computing motion using analog and binary resistive networks. *Computer*, 21, 52-63.

Koch, C., Wang, T. & Mathur, B. (1989). Computing motion in the primate's visual system. *Journal of Experimental Biology*, 146, 115-139.

Komatsu, H. & Wurtz, R. (1988). Relation of cortical areas MT and MST to pursuit eye movements. I. Localization and visual properties of neurons. *Journal of Neurophysiology*, 60(2), 580-603.

Lidén, L. (1997). A neural model of visual motion segmentation and integration. *Computational Science Graduate Fellowship Conference*, Washington D.C.

Lidén, L. (1998). A model of motion integration and segmentation in area MT. *Investigative Ophthalmology and Visual Science*, 38, 327.

Lidén, L. & Mingolla, E. (1998). Monocular occlusion cues alter the influence of terminator motion in the barber pole phenomenon. *Vision Research*, 38, 3883-3898.

Lorenceau, J. & Shiffrar, M. (1992). The influence of terminators on motion integration across space. *Vision Research*, 32(2), 263-273.

Lorenceau, J., Shiffrar, M., Wells, N. & Castet, E. (1993). Different motion sensitive units are involved in recovering the direction of moving lines. *Vision Research*, 33(9), 1207-1217.

Marcar, V., Xiao, K., Raiguel, S., Maes, H. & Orban, G. (1995). Processing of kinetically

defined boundaries in the cortical motion Area MT of the macaque monkey. *Journal of Neurophysiology*, 74(3), 1258-1270.

Marr, D. (1982). "Directional selectivity" In Marr, D. (Ed.), *Vision*, Freeman and Company, San Francisco, pp.159-182.

Marr, D. & Ullman, S. (1981). Directional selectivity and its use in early visual processing". *Proceeding of the Royal Society of London*: B, 211, 151-180.

Maunsell, J. & Van Essen, D. (1983). The connections of the middle temporal visual area (MT) and their relationship to a cortical hierarchy in the macaque monkey. *Journal of Neuroscience*, 3(12), 2563-2586.

Nakayama, K. & Loomis, J. (1974). Optical velocity patterns, velocity-sensitive neurons, and space perception: A hypothesis. *Perception*, 3, 63-80.

Nakayama, K. & Silverman, G. (1988a). The aperture problem. I. Perception of nonrigidity and motion direction in translating sinusoidal lines. *Vision Research*, 28(6), 739-746.

Nakayama, K. & Silverman, G. (1988b). The aperture problem. II. Spatial integration of velocity information along contours. *Vision Research*, 28(6), 747-753.

Nowlan, S. & Sejnowski, T. (1994). Filter selection model for motion segmentation and velocity integration. *Journal of the Optical Society of America: A*, 11(12), 3177-3200.

Pack, C. (1998) Visual cortical mechanisms of motion perception and navigation: Experiments and models. Doctoral Thesis, Boston University.

Pack, C., Grossberg, S., & Mingolla, E. (1998) Cortical processing of visual motion for smooth pursuit eye movements. *Society for Neuroscience Abstracts*, 24:3440.

Poggio, T. & Koch, C. (1985). Ill-posed problems in early vision: From computational theory to analogue networks. *Proceeding of the Royal Society of London: B*, 226, 303-323.

Poggio, T., Torre, V. & Koch, C. (1985). Computational vision and regularization theory. *Nature*, 317, 314-319.

Qian, N. & Andersen, R. (1994). Transparent motion perception as detection of unbalanced motion signals. II. Physiology. *Journal of Neuroscience*, 14(12), 7367-7380.

Qian, N., Andersen, R. & Adelson, E. (1994). Transparent motion perception as detection of unbalanced motion signals. III. Modeling. *Journal of Neuroscience*, 14(12), 7381-7392.

Raiguel, S., Van Hulle, M., Xiao, D., Marcar, V. & Orban, G. (1995). Shape and spatial distribution of receptive fields and antagonistic motion surrounds in the middle tem-

poral area (V5) of the macaque. *European Journal of Neuroscience*, 7, 2064-2082.

Ramachandran, V. & Anstis, S. (1986a). The perception of apparent motion. *Scientific American*, 254(6), 102-109.

Ramachandran, V. & Anstis, S. (1986b). Figure-ground segregation modulates apparent motion. *Vision Research*, 26(12), 1969-1975.

Reichardt, W. (1961). Autocorrelation, a principle for the evaluation of sensory information by the central nervous system. In Rosenblith, W. (Ed.), *Sensory Communication*, MIT Press, Cambridge, MA, pp.303-317.

Rodman, H. & Albright, T. (1989). Single-unit analysis of pattern-motion selective properties in the middle temporal visual area (MT). *Experimental Brain Research*, 75, 53-64.

Sachtler, W. & Zaidi, Q. (1995). Visual processing of motion boundaries. *Vision Research*, 35(6), 807-826.

Saito, H. (1993). Hierarchical neural analysis of optical flow in the macaque visual pathway. In Ono, T., Squire, L., Raichle, M., Perrett, D. & Fukuda, M. (Ed.), *Brain mechanisms of perception and memory: From neuron to Behavior*, Oxford Press, New York, pp.121-139.

Shiffrar, M. & Lorenceau, J. (1996). Increased motion linking across edges with decreased luminance contrast, edge width and duration. *Vision Research*, 36(14), 2061-2067.

Shiffrar, M. & Pavel, M. (1991). Percepts of rigid motion within and across apertures. *Journal of Experimental Psychology: Human Perception and Performance*, 17(3), 749-761.

Shimojo, S., Silverman G.H., & Nakayama, K. (1989). Occlusion and the solution to the aperture problem for motion. *Vision Research*, 29, 619-626.

Shipp, S. & Zeki, S. (1985). Segregation of pathways leading from area V2 to areas V4 and V5 of macaque monkey visual cortex. *Nature*, 315, 322-325.

Sillito, A. (1975). The contribution of inhibitory mechanisms to the receptive filed properties of neurones in the striate cortex of the cat. *Journal of Physiology*, 250, 305-329.

Snowden, R. & Braddick, O. (1989). The combination of motion signals over time. *Vision Research*, 29(11), 1621-1630.

Snowden, R. J., Treue, S., Erickson, R. G., Andersen, R. A. (1991). The response of area MT and V1 neurons to transparent motion. *Journal of Neuroscience*, 11(9), 2768-2785.

Stoner, G. & Albright, T. (1993). Image segmentation cues in motion processing: Implications for modularity in vision. *Journal of Cognitive Neuroscience*, 5(2), 129-149.

Tanaka, K., Hikosaka, K., Saito, H., Yukie, M., Fukada, Y. & Iwai, E. (1986). Analysis of local and wide-field movements in the superior temporal visual areas of the macaque monkey. *Journal of Neuroscience*, 6(1), 134-144.

Tanaka, K., Sugita, Y., Moriya, M. & Saito, H. (1993). Analysis of object motion in the ventral part of the medial superior temporal area of the macaque visual cortex. *Journal of Neurophysiology*, 69(1), 128-142.

Tse, P., Cavanagh, P. & Nakayama, K. (1998). The role of parsing in high-level motion processing. In Watanabe, T. (Ed.), Visual Motion Processing: From Computational, Neurophysiological and Psychophysical Perspectives, MIT Press, Cambridge, MA.

Vaina, L. & Grzywacz, N. (1992). Testing computational theories of motion discontinuities: A psychophysical study. *Second European Conference on Computer Vision*, Santa Margherita Ligure, Italy, 212-216.

Vaina, L., Grzywacz, N. & Kikinis, R. (1994). Segregation of computations underlying perception of motion discontinuity and coherence. *NeuroReport*, 5, 2289-2294.

Van Essen, D., Maunsell, J. & Bixby, J. (1981). The middle temporal visual area in macaque: Myeloarchitecture, connections, functional properties and topographic organization. *Journal of Comparative Neurology*, 199, 292-326.

Van Santen, J. & Sperling, G. (1984). Temporal covariance model of human motion perception. *Journal of the Optical Society of America: A*, 1(5), 451-473.

Van Santen, J. & Sperling, G. (1985). Elaborated reichardt detectors. *Journal of the Optical Society of America: A*, 2(2), 300-321.

Wallach, H. (1976). *On Perception*, Quadrangle, New York.

Wang, H., Mathur, B. & Koch, C. (1989). Computing optical flow in the primate visual system. *Neural Computation*, 1, 92-103.

Watamaniuk, S., McKee, S. & Grzywacz, N. (1995). Detecting a trajectory embedded in random-direction motion noise. *Vision Research*, 35(1), 65-77.

Watanabe, T. & Cavanagh, P. (1991). Texture and motion spreading, the aperture problem, and transparency. *Perception & Psychophysics*, 50(5), 459-464.

Watanabe, T. & Cole, R. (1995). Propagation of local motion correspondence. *Vision Research*, 35(20), 2853-2861.

Weiss, Y. & Adelson, E. (1994). Perceptually organized EM: A framework for motion seg-

mentation that combines information about form and motion. *MIT Media laboratory Perceptual Computing Section*, Tech. Report #315.

Williams, D. & Sekuler, R. (1984). Coherent global motion percepts from stochastic local motions. *Vision Research*, 24(1), 55-62.

Wurtz, R., Yamasaki, D., Duffy, C. & Roy, J. (1990). Functional specialization for visual motion processing in primate cerebral cortex. In (Ed.), *Cold Spring harbor Symposia on Quantitative Biology*, Cold Spring Harbor Laboratory Press, pp.717-727.

Yuille, A. L. & Grzywacz (1988). A computational theory for the perception of coherent visual motion. *Nature*, 333(6168), 71-74.

Zeki, S. (1974a). Cells responding to changing image size and disparity in the cortex of rhesus monkey. *Journal of Physiology*, 242, 827-841.

Zeki, S. (1974b). Functional organization of a visual area in the posterior bank of the superior temporal sulcus of the rhesus monkey. *Journal of Physiology*, 236, 549-573.

Appendix

This section describes the equations used to implement the model discussed in this paper. The first three sections describe the calculations used for the input nodes (LMD), the integration nodes, (*I*) and the segmentation nodes (*S*) respectively. For convenience, a summary of all variables and their functions in the model is provided at the end. The model was implemented in the C programming language, and all simulations were run on a Silicon Graphics workstation running the IRIX operating system. Equations were numerically integrated using floating point values. For all equations listed below, position is represented by superscripts whereas directional preference is represented by subscripts.

Appendix 1 - Input Nodes (LMDs)

The activity of input nodes is calculated using a weighted correlational scheme. For each spatial location there are eight LMD cells, each corresponding to one direction of motion. For each position in space, the activity of LMD cells is calculated by comparing a weighted window of grey-level pixel values around that position to eight different shifts of the same window at later points in time. The activity of a given LMD cell, ν with position subscripts, x, y, and directional preference superscript, n, is given by:

$$\boldsymbol{v}_{x,y}^{n} = \left(1 - \sum_{(i,j) \in K} k_{i,j} \left| \boldsymbol{I}(t)_{(x+i),(y+j)} - \boldsymbol{I}(t+1)_{\left(x+i+\gamma_i^n\right),\left(y+j+\gamma_j^n\right)} \right| \right) + \left(1 - \sum_{(i,j) \in K} k_{i,j} \left| \boldsymbol{I}(t)_{(x+i),(y+j)} - \boldsymbol{I}(t+2)_{\left(x+i+\gamma_i^n\right),\left(y+j+\gamma_j^n\right)} \right| \right) \quad \text{(EQ 1)}$$

$\boldsymbol{I}(t)_{x,y}$ corresponds to the grey-level pixel intensity in the input image at position x, y and time t. K is the spatial extent of the correlation kernel. γ_i^n, γ_j^n are the x and y positional shifts for each point within the kernel for direction n. $k_{i,j}$ is the roughly Gaussian weight matrix corresponding to the correlation kernel. The weights were normalized such that:

$$\sum_{(i,j)\in K} k_{i,j} = 1 . \qquad \text{(EQ 2)}$$

When perfect correlation occurs for a shift in a specific direction, the grey-scale values of one window matches the grey-scale values of the window shifted in that direction at a later time step. The summation value is therefore 0. On the other hand, when perfect anti-correlation occurs, the weighted sum is 1, since the weight matrix is normalized. By subtracting the summation from a value of one, the result is assured to be 1 when perfect correlation occurs and 0 when perfect anti-correlation occurs.

Two points in time are used to capture the fact that the multiple speeds may be present in the input stimulus. A more complete model would use multiple sets of LMD cells each tuned to different speeds, rather than the simple summation employed here.

Appendix 2 - Integration (*I*) Cells

The dynamics of the integration system are implemented by a single equation which incorporates all three computational principles: inhibition within spatial location, Γ, short range excitation, Λ and long range inhibition, Ω, as well as inhibition from the segmenta-

tion nodes, χ (See Figure 2). The activity of an integration node, with position subscripts, x, y and direction selectivity superscript n is given by:

$$\frac{d}{dt}\boldsymbol{m}_{xy}^{n} = \left(1 - \boldsymbol{m}_{xy}^{n}\right)\left(B\nu_{xy}^{n} + E\Lambda_{xy}^{n}\right) - \left(\boldsymbol{m}_{xy}^{n}\right)\left(D_{m} + C\Gamma_{xy}^{n} + A\Omega_{xy}^{n} + F\chi_{xy}\right) . \quad \text{(EQ 3)}$$

The equation consists of excitatory and inhibitory terms gated by the node's own activity. This keeps the activation of each integration node in the range [0,1]. Input from LMDs is denoted ν and weighted by parameter B. Each node also has a decay rate D_m. The three computational processes, Γ, Λ and Ω are weighted by parameters E, C and A respectively and are given by:

(1) *Inhibition within spatial location across direction,* Γ, with an inter-directional connectivity matrix α where k and n are direction indices, given by

$$\left(\Gamma_{xy}^{n} = \sum_{k=1}^{\sigma} \boldsymbol{m}_{xy}^{k}\alpha^{kn}\right) . \quad \text{(EQ 4)}$$

(2) *Short range excitation,* Λ, with an inter-directional connectivity matrix μ_1 and an inter-spatial connectivity matrix μ_2 with direction indices k and n, and positional indices x, y, i and j, given by

$$\Lambda_{xy}^{n} = \sum_{(i,j) \in \lambda}\left(\sum_{k=1}^{\sigma} \boldsymbol{m}_{ij}^{k}\mu_{1}^{kn}\right)\mu_{2}^{ijxy} . \quad \text{(EQ 5)}$$

(3) *Long range inhibition,* Ω, With an inter-directional connectivity matrix ϕ_1 and an

inter-spatial connectivity matrix ϕ_2, given by

$$\left(\Omega_{xy}^{n} = \sum_{(i,j) \in \omega} \left(\sum_{k=1}^{\sigma} \boldsymbol{m}_{ij}^{k} \phi_{1d}^{kn} \right) \phi_{2}^{ijxy} \right) . \quad \text{(EQ 6)}$$

All inter-directional connectivity matrices, α, μ_1 and ϕ_1 are roughly Gaussian in shape.

Inhibition by the segmentation nodes, χ is weighted by parameter F and is simply the sum of all segmentation node activity at that position:

$$\chi = \sum_{n=1}^{\sigma} \boldsymbol{s}_{xy}^{n} , \quad \text{(EQ 7)}$$

where $\boldsymbol{s}$ is a segmentation node at position x, y with directional tuning n.

Appendix 3 - Segmentation (*S*) Cells

For each position in space there is a set of eight segmentation nodes, s with position i, j and directional selectivity n. Ψ, the lateral excitation from other segmentation nodes, is defined as:

$$\Psi_{xy}^{n} = \sum_{(i,j) \in \psi} \boldsymbol{s}_{ij}^{n} , \quad \text{(EQ 8)}$$

where ψ is the spatial extent of lateral excitation. χ, the input from LMDs, is defined as:

$$\chi_{xy} = \sum_{n=1}^{\sigma} \nu_{xy}^{n} , \quad \text{(EQ 9)}$$

where ν_{xy}^{n} is the LMD input at spatial position x, y for direction n. The summation

results in directionally non-specific input. Finally, the surround inhibition, $\Im$, is defined as:

$$\Im_{xy}^{n} = \sum_{(i,j)\in\varepsilon} \boldsymbol{m}_{ij}^{n}, \qquad \text{(EQ 10)}$$

where ε is the spatial extent of the surround inhibition. Thus a segmentation cell is inhibited when integration cells in the surround with the same directional tuning are active. In other words, the segmentation cell is suppressed in the presence of whole field movements.

The activity of a segmentation node $\boldsymbol{s}$ with directional preference n at position x, y is given by:

$$\frac{d}{dt}\boldsymbol{s}_{xy}^{n} = \left(1 - \boldsymbol{s}_{xy}^{n}\right)\left(G\Theta_{xy}^{n}\right) - \left(\boldsymbol{s}_{xy}^{n}\right)(D_s + H\Im_{xy}^{n}), \qquad \text{(EQ 11)}$$

where D_s is the decay rate. Θ is the gated excitation of the segmentation nodes weighted by parameter G, and $\Im$ is the inhibition of the segmentation nodes weighted by parameter H.

Each segmentation cell's input is gated (Figure 3), such that its input is only effective if the cell also receives lateral excitation from either segmentation nodes or input from LMDs. The gated input Θ is given by:

$$\Theta_{xy}^{n} = \begin{cases} \Upsilon_{xy}^{n} & \text{if} \quad \Psi_{xy}^{n} > 0 \text{ or } \chi > 0 \\ 0 & \text{otherwise} \end{cases}, \qquad \text{(EQ 12)}$$

where Υ is the total excitatory input, Ψ is the lateral excitation, and χ is the input from LMDs. The total excitatory activity, Υ, is given by:

$$\Upsilon_{xy}^{n} = \sum_{(i,j) \in \theta_c} \left[\boldsymbol{m}_{ij}^{n} - J \sum_{d \neq n} \boldsymbol{m}_{ij}^{d} \right]^{+} + \sum_{(i,j) \in \theta_s} \boldsymbol{m}_{ij}^{\bar{n}} \quad . \qquad \text{(EQ 13)}$$

The first term sums over the kernel center, θ_c. Integration cells, $\boldsymbol{m}$, within the center that possess the same directional preferences, n, are excitatory. Those possessing all other directional preferences, $d \neq n$, are inhibitory and weighted by a parameter, J. $[\ ]^{+}$ refers to a half-wave rectification that ensures that the sum is always positive. In this way the segmentation cells receive more excitation from disambiguated areas of visual space (i.e. locations in which fewer directions are active at the integration cell level). The second term sums over the receptive field surround, θ_s. In the surround, integration cells, $\boldsymbol{m}$, of the opposite directional preference, $\bar{n}$, are excitatory. The conditional statement, equation 12, ensures that the excitatory input, Υ, is only effective when either lateral excitation, Ψ or input from V1, χ is also present.

Summary of Variables

Network Components

m An integration node (I)

s A segmentation node (S)

v An input node (LMD)

I A pixel intensity (input image)

Indices

n, d - Directional preference

t - Time

x, i - Horizontal spatial position

y, j - Vertical spatial position

Equation Variables

Λ Short range excitation for *integration cells*

Γ Inhibition within location for *integration cells*

Ω Long range inhibition for *integration cells*

Θ Gated center-surround excitation for *segmentation cells*

Υ Center-surround excitation for *segmentation cells*

$\Im$ Center-surround inhibition for *segmentation cells*

Ψ Lateral segmentation cell input for *segmentation cells*

Spatial Dimensions

γ_i^n Correlation shift in x component for direction preference n for *input cells*

γ_j^n Correlation shift in y component for direction preference n for *input cells*

K Spatial extent of correlation kernel of *input cells*

k Weight of a position within the correlation kernel of *input cells*

α Connectivity matrix for inhibition within location (Γ) of *integration cells*

μ_1 Inter-directional connect matrix for short range excitation (Λ) of *integration cells*

μ_2 Inter-spatial connectivity matrix for short range excitation (Λ) of *integration cells*

ϕ_1 Inter-directional connect matrix for long range inhibition (Ω) of *integration cells*

ϕ_2 Inter-spatial connectivity matrix for long range inhibition (Ω) of *integration cells*

λ Spatial extent of short range excitation (Λ) for *integration cells*

ω Spatial extent of long range inhibition (Ω) for *integration cells*

ψ Spatial extent of lateral excitation for *segmentation cells*

θ_c Spatial extent of center for center-surround input for *segmentation cells*

θ_s Spatial extent of surround for center-surround input for *segmentation cells*

ε Spatial extent of surround inhibition for *segmentation cells*

Adjustable Parameters

A Weight parameter for long range inhibition (Ω) of *integration cells*

B Weight parameter for LMD input ($\mathbf{v}$) of *integration cells*

C Weight parameter for inhibition within spatial location (Γ) of *integration cells*

E Weight parameter for short range excitation (Λ) of *integration cells*

F Weight parameter for segmentation cell inhibition (χ) of *integration cells*

D_m Decay rate parameter for *integration cells*

D_s Decay rate parameter for *segmentation cells*

G Weight parameter for center surround excitation (Θ) of *segmentation cells*

H Weight parameter for center surround inhibition ($\Im$) of *segmentation cells*

J Weight parameter of center surround center inhibition (Υ) of *segmentation cells*

Summary of Parameters (same values used for all simulations)

$A = 1.0,\ B = 0.8,\ C = 6,\ D_m = 0.5,\ D_s = 0.6,\ E = 0.7,\ F = 0.5,\ G = 1.7,$

$H = 1.2,\ J = 0.4,\ \alpha, \phi_1 = \left\{ \begin{matrix} 0.9^{n,\,n-3}, 0.5^{n,\,n-2}, 0.5^{n,\,n-1}, 0.0^{n,\,n} \\ 0.5^{n,\,n+1}, 0.5^{n,\,n+2}, 0.9^{n,\,n+3}, 1.0^{n,\,n+4} \end{matrix} \right\}$

As the kernel, λ consisted of only the most adjacent surrounding units, the inter-spatial connectivity matrix, μ_2 was simply 1.0 for all values. Were larger spatial scales employed the connectivity matrix would need a gaussian fall-off with increasing distance. The kernel, ω, consisted of surrounding units of a distance of three units or less. ϕ_2, the inter-spatial connectivity matrix was 0.5 for the most adjacent spatial position, 1.0 for an offset of one, and 0.5 for an offset of two spatial positions.