Laminar cortical dynamics of visual form and motion interactions during coherent object motion perception

by

J. Berzhanskaya[*1], S. Grossberg[+2], E. Mingolla[#2]

[1]Krasnow Institute for Advanced Study
George Mason University
Fairfax, VA, 22030
[2] Department of Cognitive and Neural Systems
Center for Adaptive Systems
and
Center of Excellence for Learning in Education, Science, and Technology
Boston University
677 Beacon Street
Boston, MA 02215

Submitted: May, 2006
Revised: January, 2007

Corresponding Author: S. Grossberg, steve@bu.edu

Abstract

How do visual form and motion processes cooperate to compute object motion when each process separately is insufficient? Consider, for example, a deer moving behind a bush. Here the partially occluded fragments of motion signals available to an observer must be coherently grouped into the motion of a single object. A 3D FORMOTION model comprises five important functional interactions involving the brain's form and motion systems that address such situations. Because the model's stages are analogous to areas of the primate visual system, we refer to the stages by corresponding anatomical names. In one of these functional interactions, 3D boundary representations, in which figures are separated from their backgrounds, are formed in cortical area V2. These depth-selective V2 boundaries select motion signals at the appropriate depths in MT via V2-to-MT signals. In another, motion signals in MT disambiguate locally incomplete or ambiguous boundary signals in V2 via MT-to-V1-to-V2 feedback. The third functional property concerns resolution of the aperture problem along straight moving contours by propagating the influence of unambiguous motion signals generated at contour terminators or corners. Here, sparse "feature tracking signals" from, e.g., line ends, are amplified to overwhelm numerically superior ambiguous motion signals along line segment interiors. In the fourth, a spatially anisotropic motion grouping process takes place across perceptual space via MT-MST feedback to integrate veridical feature-tracking and ambiguous motion signals to determine a global object motion percept. The fifth property uses the MT-MST feedback loop to convey an attentional priming signal from higher brain areas back to V1 and V2. The model's use of mechanisms such as divisive normalization, endstopping, cross-orientation inhibition, and long-range cooperation is described. Simulated data include: the degree of motion coherence of rotating shapes observed through apertures, the coherent vs. element motion percepts separated in depth during the chopsticks illusion, and the rigid vs. non-rigid appearance of rotating ellipses.

Keywords: motion perception, depth perception, perceptual grouping, prestriate cortex, V1, V2, MT, MST

Introduction. Visual motion perception requires the solution of two complementary problems of *motion integration* and *motion segmentation* (Braddick, 1993). Motion integration joins nearby signals into a single percept of object motion, while segmentation keeps motion signals separate as belonging to different objects. These problems become particularly acute when an object moves behind multiple occluders. Then the various object parts are segmented by the occluders, but the visual system can often integrate these parts into a percept of coherent object motion. Studying conditions under which the visual system can and cannot accomplish correct segmentation and integration provides important cues to the processes that are used by the visual system to create object motion percepts during normal viewing conditions.

The present article further develops a 3D FORMOTION model, components of which were introduced by Baloch and Grossberg (1997), Chey, Grossberg, and Mingolla (1997, 1998), Francis and Grossberg (1996), and Grossberg, Mingolla, and Viswanathan (2001). The model explains some challenging percepts during which small changes in object or contextual cues can dramatically change motion percepts from integration to segmentation. As the model's name suggests, it proposes how form and motion processes interact to form coherent percepts of object motion in depth. The present work focuses on the following form-motion (or *formotion*) binding issues: How do form-based 3D figure-ground separation mechanisms in cortical area V2 interact with directionally selective motion grouping mechanisms in cortical areas MT and MST to preferentially bind together some motion signals more easily than others? In cases where form-based figure-ground mechanisms are insufficient, how do motion and attentional cues from cortical area MT facilitate figure-ground separation within cortical area V2 via MT-to-VI-to-V2 feedback? Finally, how does the global organization of the motion direction field in areas MT and MST influence whether the percept of an object's form looks rigid or deformable through time?

The model goes beyond earlier motion models both by introducing novel formotion binding mechanisms and by proposing how *laminar* cortical circuits realize these mechanisms. These circuits embody explicit predictions about the functional roles that are played by the corresponding cells in the brain. The model extends to the motion system a program of developing laminar models of cortical circuits that has already explained many perceptual and brain data about 3D form perception in cortical areas V1, V2, and V4 (Grossberg, 1999; Cao and Grossberg, 2005; Grossberg and Howe, 2003; Grossberg, Mingolla, and Ross, 1997; Grossberg and Raizada, 2000; Grossberg and Seitz, 2003; Grossberg and Swaminathan, 2004; Grossberg and Williamson, 2001; Grossberg and Yazdanbakhsh, 2005; Raizada and Grossberg, 2003), as well as about cognitive working memory, sequence learning, and variable-rate sequential performance (Grossberg and Pearson, 2006).

The model proposes solutions to several basic problems of motion perception, including the *aperture problem*. Wallach (1935/1996) first showed that the motion of a featureless line seen behind a circular aperture is perceptually ambiguous: no matter what may be the real direction of motion, the perceived direction is perpendicular to the orientation of the line; *i.e.*, the *normal component* of motion. The aperture problem is faced by any localized neural motion sensor, such as a neuron in the early visual pathway, which responds to a local contour moving through an aperture-like receptive field. In contrast, a moving dot, line end or corner provides unambiguous information about an object's true motion direction (Shimojo, Silverman and Nakayama, 1989). The model proposes how such moving visual features activate cells in the brain that compute *feature-tracking* signals which can disambiguate an object's true direction of motion.

A key issue concerns the assignment of motion to an object boundary when motion integration interpolates two contiguous parts of a scene, since not all line ends signal motion of an object correctly. In the example in Figure 1, motion of the left line end corresponds to the real motion of the line. The right line end is formed by the boundary between the line and a stationary occluder, and its motion provides little information about the motion of the line. This issue has been in the vision literature for a long time; e.g., see Bregman (1981) and Kanizsa (1979). Nakayama, Shimojo and Silverman (1989) have suggested classification of terminators as *intrinsic* and *extrinsic*: an intrinsic terminator belongs to the moving object; an extrinsic one belongs to the occluder. Motion of intrinsic terminators is taken into account in computing the motion direction of an object, while motion of extrinsic terminators is generally ignored (Shimojo et al., 1989; Duncan, Albright and Stoner, 2000). Lidén and Mingolla (1998), however, showed that the influence of extrinsic terminators on direction of perceived motion in a barberpole display that contains occluding surfaces was reduced, rather than abolished, as compared to comparable intrinsic terminators.

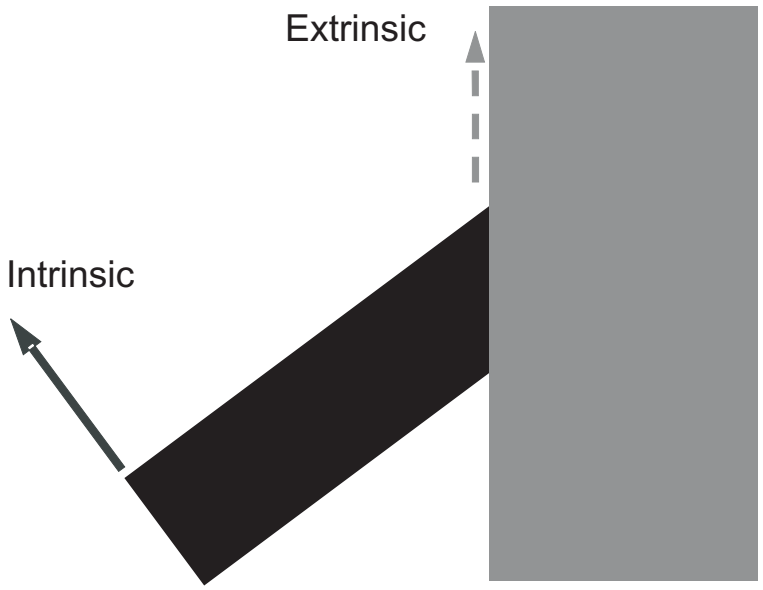

Figure 1. Extrinsic and intrinsic terminator motions are different. The local motion of the intrinsic terminator on the left reflects the true object motion, while the local motion of extrinsic terminator traces the vertical outline of the occluder.

The FACADE model (Grossberg, 1994, 1997; Kelly and Grossberg, 2000) of 3D form vision and figure-ground separation proposed how boundaries in 2D images are assigned to different objects in different depth planes, and thereby offered a mechanistic explanation of the properties of extrinsic and intrinsic terminators. A precursor of the present 3D FORMOTION model (Grossberg, Mingolla, and Viswanathan, 2001) proposed how FACADE figure-ground separation in cortical area V2, combined with formotion interactions from area V2 to MT, enable intrinsic terminators to create strong motion signals on a moving object, while extrinsic terminators create weak ones.

Simulations by Grossberg *et al.* (2001) assumed that figure-ground separation had already occurred within the form system and used depth-separated boundaries from V2 as inputs to the motion system. The present model starts with motion signals in V1, where the separation in depth has not yet occurred, and predicts how V2-to-MT boundary signals can selectively support V1-to-MT motion signals at the correct depths, while suppressing motion signals at the same visual location but at different depths.

The prediction that V2-to-MT signals can capture motion signals at a given depth reflects the hypothesis that the form and motion streams compute *complementary* properties (Grossberg, 1991, 2000): the V1-V2 cortical stream, acting alone, is predicted to compute precise oriented

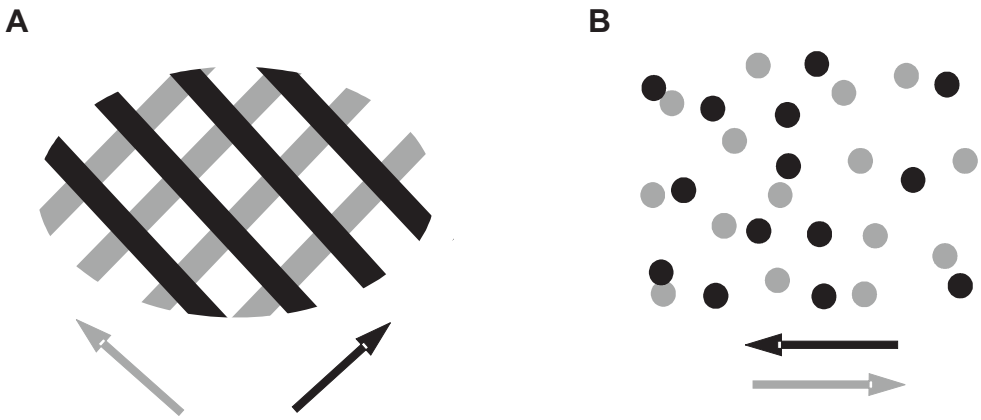

Figure 2. Plaids and transparent motion. Grayscale is added for illustration purposes only. (A) Overlapping gratings under certain conditions can cohere. Under other conditions, they can separate and be perceived as sliding over each other in the directions perpendicular to the gratings (arrows). (B) Similar effects can be observed with two sheets of randomly positioned dots moving in two different directions**.**

depth estimates (indeed, 3D boundary representations), but coarse directional motion signals, whereas the V1-MT cortical stream computes coarse oriented depth estimates, but precise directional motion estimates. The 3D boundary representations that are computed in V2 are predicted to overcome these complementary deficiencies within the form and motion streams. This is predicted to occur via V2-to-MT inter-stream interactions, called *formotion* interactions, which use signals from V2 to capture motion signals in MT to lie at the correct depths. In this way, precise form-and-motion-in-depth estimates are achieved in MT, which can be used to generate good target tracking estimates. Ponce, Lomber, and Born (2006) have recently reported neurophysiological data that are consistent with the prediction that V2 imparts finer disparity sensitivity onto MT.

The V2-to-MT motion selection mechanism clarifies why we tend to perceive motion of visible objects and background features, but not of the intervening empty spaces between them. This may not seem to be a serious problem if we just consider the motion signals of which we are consciously aware. However, when one considers how motion signals can have an influence on visible features across empty space, as during induced motion, without causing visible motion within the intervening space that is devoid of visible features, one readily sees that it is a phenomenon that requires explanation (Duncker, 1929/1937). Motion selection in MT using depth-separated form boundaries from V2 is, we believe, a part of the explanation, since only those motion signals in MT that are captured by a V2 form boundary can be used to form percepts when such boundaries are active

These V2-to-MT formotion signals overcome one sort of uncertainty in cortical computation. Another sort of uncertainty is overcome via MT-to-V1 feedback signals which can help to separate boundaries in V1 and V2 where they cross in feature-absent regions (cf. the chopsticks illusion below) using motion signals from MT.

Another factor that influences motion perception is adaptation. Motion signals at the positions of a static extrinsic terminator in can adapt, and therefore attenuate. Moving intrinsic terminators, on the other hand, generate strong motion signals. As local motion signal direction

and strength are computed, a motion integration process in MT-MST decides the winning motion direction in the case of a single moving line, as in Figure 1.

What happens if multiple moving objects overlap? Experiments on plaids and random dot motion have demonstrated at least two possible perceptual outcomes (Ferrera and Wilson, 1987, 1990; Kim and Wilson, 1993; Snowden et al., 1991; Stoner and Albright, 1998; Stoner, Albright, and Ramachandran, 1990; Trueswell and Hayhoe, 1993). See Figure 2. First, a display can separate into two depth planes, forming a transparent motion percept, where two dot-filled planes or two gratings slide one over another. Second, if the directions of motions are compatible, then displays can produce a percept of coherent motion of a unified pattern, and no separation in depth occurs. Under prolonged viewing, the same display can perceptually alternate between coherent plaid motion and different motions separated in depth (Hupé and Rubin, 2003). Our present work focuses on the distinction between the two types of motion that are generally obtained, on a shorter time scale, with exposures of up to a few seconds. An earlier version of the present model (Grossberg, Mingolla, and Viswanathan, 2001, Section 3.10) discussed how adaptation can influence percepts of coherent and incoherent plaid motion.

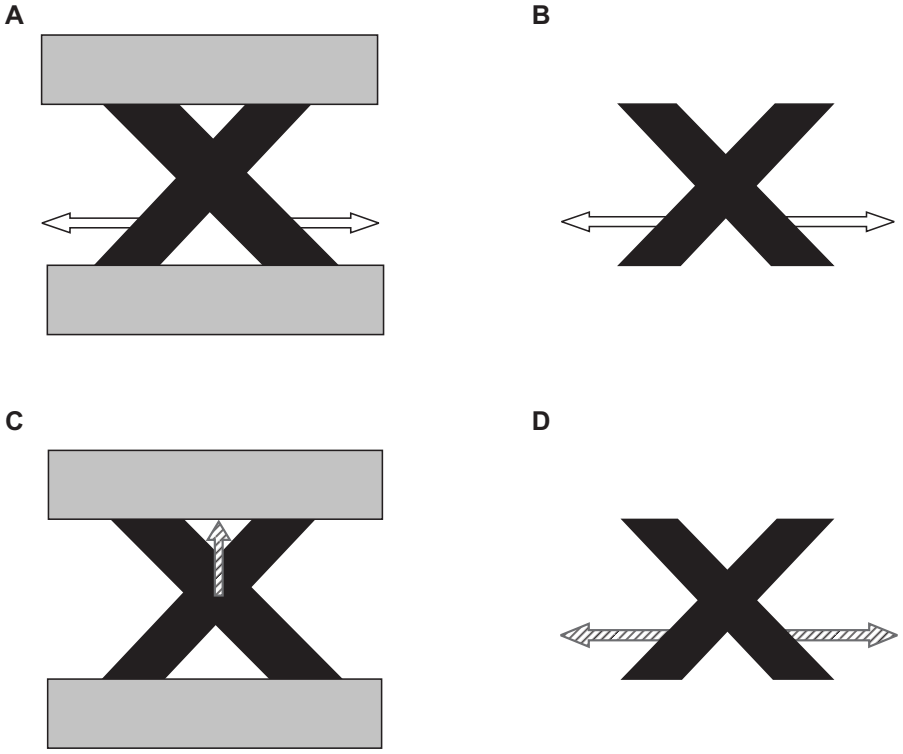

Figure 3. Chopsticks Illusion. Actual chopsticks motion (clear arrows, top) is equivalent in (A) and (B), with visible and invisible occluders, respectively. Visible occluders result in a coherent vertical motion percept (C, hatched arrow). Invisible occluders result in the percept of two chopsticks sliding in opposite directions (D).

As noted above, while separation in depth can happen purely in the motion system, occluder information from the form system can modulate the calculation of motion signals (Stoner and Albright, 1996, 1998). For example, the present article models the motion percepts that are

generated by a chopsticks display (Anstis, 1990). See Figure 3. The bars in this display undergo translational motion, and may be thought of as a simplified plaid motion display. When the chopsticks move horizontally, their intersection moves vertically. In the case of visible occluders (Figure 3A), the intersection motion prevails and vertical motion of a single X-shaped object is perceived. In the case where the chopstick ends are visible (Figure 3B) — that is, the occluder is invisible — the percept is of two chopsticks moving in opposite horizontal directions and separated in depth. This depth separation cannot happen based only on the boundaries of the X-shaped form, since the boundaries near the middle of the X do not complete either bar explicitly. The 3D FORMOTION model proposes how signals from the motion to the form stream via MT-to-V1 feedback can initiate the process whereby these ambiguous boundaries can be completed and separated in depth within the form stream.

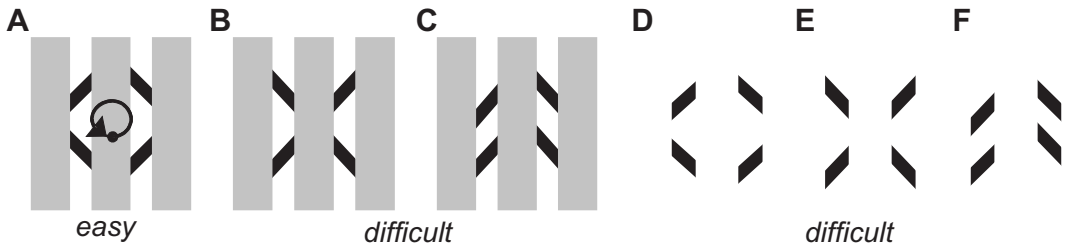

Figure 4. Snapshots of the Lorenceau-Alais displays. Visible (A-C) and invisible (D-F) occluder cases. See text for details.

Often the shape of a moving object is more complex than that of a line, and can affect the outcome of motion integration. The present article simulates data of Lorenceau and Alais (2001), who studied different shapes moving in a circular-parallel motion behind occluders (Figure 4). Observers had to determine the direction of motion, clockwise or counterclockwise. The percent of correct responses depended on the type of shape, and on the visibility of the occluders. In the case of a diamond (Figure 4A), a single, coherent, circular motion of a partially occluded rectangular frame was easy to perceive across the apertures. In the case of an arrow (Figure 4C), two objects with parallel sides were seen to generate out-of-phase vertical motion signals in adjacent apertures. Local motion signals were identical in both displays, and only their spatial arrangement differed. Alais and Lorenceau suggested that certain shapes (such as arrows) "veto" motion integration across the display, while others (such as diamond) allow it.

The 3D FORMOTION model explains the data without using a veto process. The model proposes that the motion grouping process uses anisotropic direction-sensitive receptive fields that preferentially integrate motion signals within a given direction across gaps produced by the occluders. The explanation of Figures 4D-F follows in a similar way, with the additional factor that the ends of the bars possess intrinsic terminators that can strongly influence the perceived motion direction of the individual bars.

Another example of where percepts of rotational motion involve motion grouping is the "gelatinous ellipses" display (Vallortigara et al., 1988, Weiss and Adelson, 2000). See Figure 5. When the "thin" (high aspect ratio) and the "thick" (low aspect ratio) ellipses rotate around their centers, the perception of their shapes is strikingly different. The thin ellipse is perceived as a rigid rotating form, whereas the thick one is perceived as deforming non-rigidly through time. Here, the differences in 2D geometry result in differences of the spatiotemporal distribution of motion direction signals that are grouped together through time. When these motion signals are

consistent with the coherent motion of a single object, then the motion grouping generates a percept of a rigid rotation. When the motion field decomposes after grouping into multiple parts, with motion trajectories incompatible with a rigid form, a non-rigid percept is obtained. Motion-to-form projections from MT to V1 can once again help to explain these distinct outcomes. The ability of nearby "satellites" to convert the non-rigid percept into a rigid one can also be explained by motion grouping. Weiss and Adelson (2000) have proposed that such a percept can be explained via a global optimization process. Motion grouping provides a biologically plausible alternative proposal.

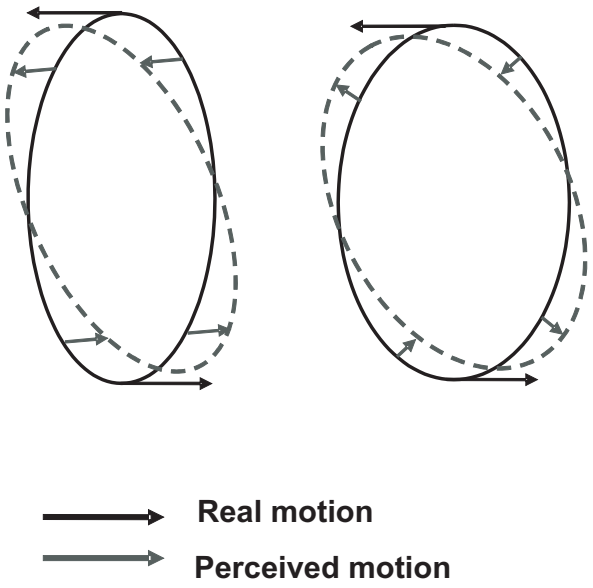

Figure 5. Rotating ellipses.
Rigid (left) and nonrigid (right) percepts.

In summary, all of the data considered here illustrate how the brain may use both form and motion information, and their interaction, to derive a global percept of object motion. Form and motion processes, such as those in V2/V4 and MT/MST, occur in the What ventral and Where dorsal cortical processing streams. As noted above, related modeling work has proposed that key mechanisms within the What and Where streams obey computationally *complementary* laws (Grossberg, 2000): The ability of each process to compute some properties prevents it from computing other, complementary, properties. Examples of such complementary properties include boundary completion vs. surface filling-in—within the (V1 interblob)-(V2 interstripe) and (V1 blob)-(V2 thin stripe) streams, respectively—and, more relevant to the results herein, boundary orientation vs. motion direction, and fine boundary disparity vs motion direction—within the V1-V2 and V1-MT streams, respectively. The present article clarifies some of the interactions between form and motion processes that enable them to overcome their complementary deficiencies and to thereby compute more informative representations of unambiguous object motion.

In our simulations, each model layer consists of a 60x60 matrix with multiple cells that code for different properties such as line orientation or motion direction at each position. A detailed model description is provided after simulations are presented in Appendix A. The 3D FORMOTION model comprises five important functional interactions involving the brain's form and motion systems that allow it to perform appropriate grouping and segmentation of fragmentary motion signals caused by occlusion of objects intervening between the viewer and a moving object. Because the model's stages are analogous to areas of the primate visual system, we refer to the stages by corresponding anatomical names. In one of these functional interactions, 3D boundary representations, in which figures are separated from their backgrounds, are formed in cortical area V2. These depth-selective V2 boundaries select motion signals at the appropriate depths in MT via V2-to-MT signals. In another, motion signals in MT disambiguate locally incomplete or ambiguous boundary signals in V2 via MT-to-V1-to-V2

feedback. The third functional property concerns resolution of the aperture problem along straight moving contours through appeal to unambiguous motion signals generated at contour terminators or corners. Here, sparse "feature tracking signals" from, e.g., line ends, are amplified to overwhelm numerically superior ambiguous motion signals along line segment interiors. In the fourth, a spatially anisotropic motion grouping process propagates across perceptual space via MT-MST feedback to integrate veridical feature-tracking and ambiguous motion signals to determine a global object motion percept. The fifth property is the capacity of the MT-MST feedback loop to convey an attentional priming signal from higher brain areas back to V1 and V2.

3D FORMOTION Model

The main components of the 3D FORMOTION model are a *form processing stream* and a *motion processing stream*. These streams interact in specific ways, as indicated in Figures 6 and 7.

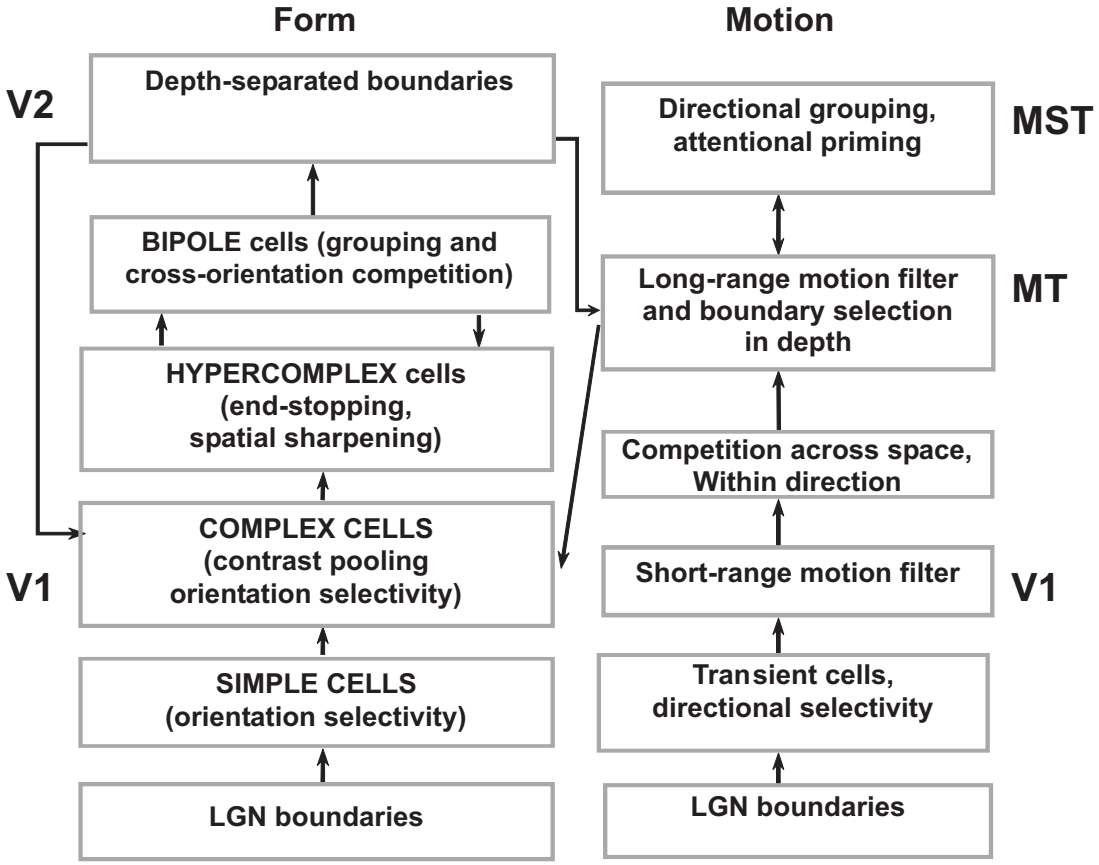

Figure 6. Schematic view of the 3D FORMOTION model. See text for details.

The Form Processing System

The model's form processing system comprises six stages, as shown on the left sides of Figures 6 and 7. Input to the model is represented by distinct ON and OFF cells, whose properties derive from on-center off-surround and off-center on-surround network interactions, similar to those demonstrated by LGN cells. Because of our use of simple black and white images, retinal and

LGN processes of both the form and motion streams can be treated as a simplified lumped processing stage. Subsequent processing in the form stream includes simple cells for initial registration of boundary orientations, followed by complex and hypercomplex stages that perform: (a) pooling across simple cells tuned to opposite contrast polarities; (b) divisive normalization that reduces the amplitude of multiple ambiguous orientations in a region, (3) end-stopping that enhances activity at line-ends, and (4) spatial sharpening that prevents excessive blurring of boundary localization. Long-range bipole cells, indicated by the figure-8 shape in Figure 7, act like statistical "and" gates that group approximately collinear boundary signals. Grouping is followed by a stage of cross-orientation competition that reinforces boundary signals with superior support from neighboring boundaries at the expense of spatially overlapping signals of non-preferred orientations. Finally, an assignment of boundaries into one or both of two simulated depth representations is accomplished, as is next described.

Perceptual Grouping and Figure-Ground Separation of 3D Form. The FACADE boundary completion process is called the Boundary Contour System, or BCS (Figures 6 and 7, left). The BCS predicts how boundaries of occluding surfaces are separated from occluded surfaces in depth, including the separation of extrinsic vs. intrinsic boundaries (Grossberg, 1994, 1997; Grossberg and Yazdanbakhsh, 2005; Kelly and Grossberg, 2000), within the pale stripes of V2. One cue of occlusion in a 2D image is a T-junction. The black bar in Figure 8A forms a T-junction with the gray bar (Figure 8B). The top of the T belongs to the occluding black bar, while the stem belongs to the occluded gray bar. Bipole long-range grouping (Figure 8C) strengthens the horizontal boundary, while short-range competition weakens the vertical boundary (Figure 8D). This end gap in the vertical boundary initiates the process of separating occluding and occluded boundaries. In other words, basic properties of perceptual grouping are predicted to initiate the separation of figures from their backgrounds, without the use of explicit T-junction operators. Such figure-ground separation is a crucial competence of the 3D FORMOTION model. It enables the model to distinguish extrinsic from intrinsic terminators, and to thereby compute appropriate signals in the motion stream, as will be explained when that part of the model is described.

In order to simplify our simulations, the 3D FORMOTION model does not include all the stages of boundary and surface interaction that complete figure-ground separation. That these mechanisms work has been demonstrated elsewhere (Fang and Grossberg, 2004; Grossberg and Yazdanbaksh, 2005; Kelly and Grossberg, 2000). Instead, as soon as T-junctions have been detected by the model dynamical equations, boundaries are algorithmically separated in depth. That is, the representation of boundaries is assigned by our simulation code to the depth where the boundary would be represented if a "full-blown" FACADE simulation were done. In particular, static occluders are assigned to the near depth and lines with extrinsic terminators are assigned to the far depth. At a T-junction, the horizontal boundary will be represented in Depth 1 and the vertical boundary in Depth 2. Because of this computational shortcut, thin idealized boundaries, positioned at the same locations as input boundaries are used to select motion signals via V2-MT projections (see Appendix A). The effect of motion on boundary position shifts is not considered here, but was explored in simulations of flash-lag and flash-drag effects by Berzhanskaya, Grossberg and Mingolla (2004). V2 boundaries are used to provide both V2-to-MT motion selection signals (Equation A14) and V2-to-V1 depth-biasing feedback (Equation A28) (Figure 7, top-left). While V2-to-V1 feedback is orientation-specific, the V2-to-MT projection sums boundary signals over all orientations, just as motion signals do at MT (Albright, 1984).

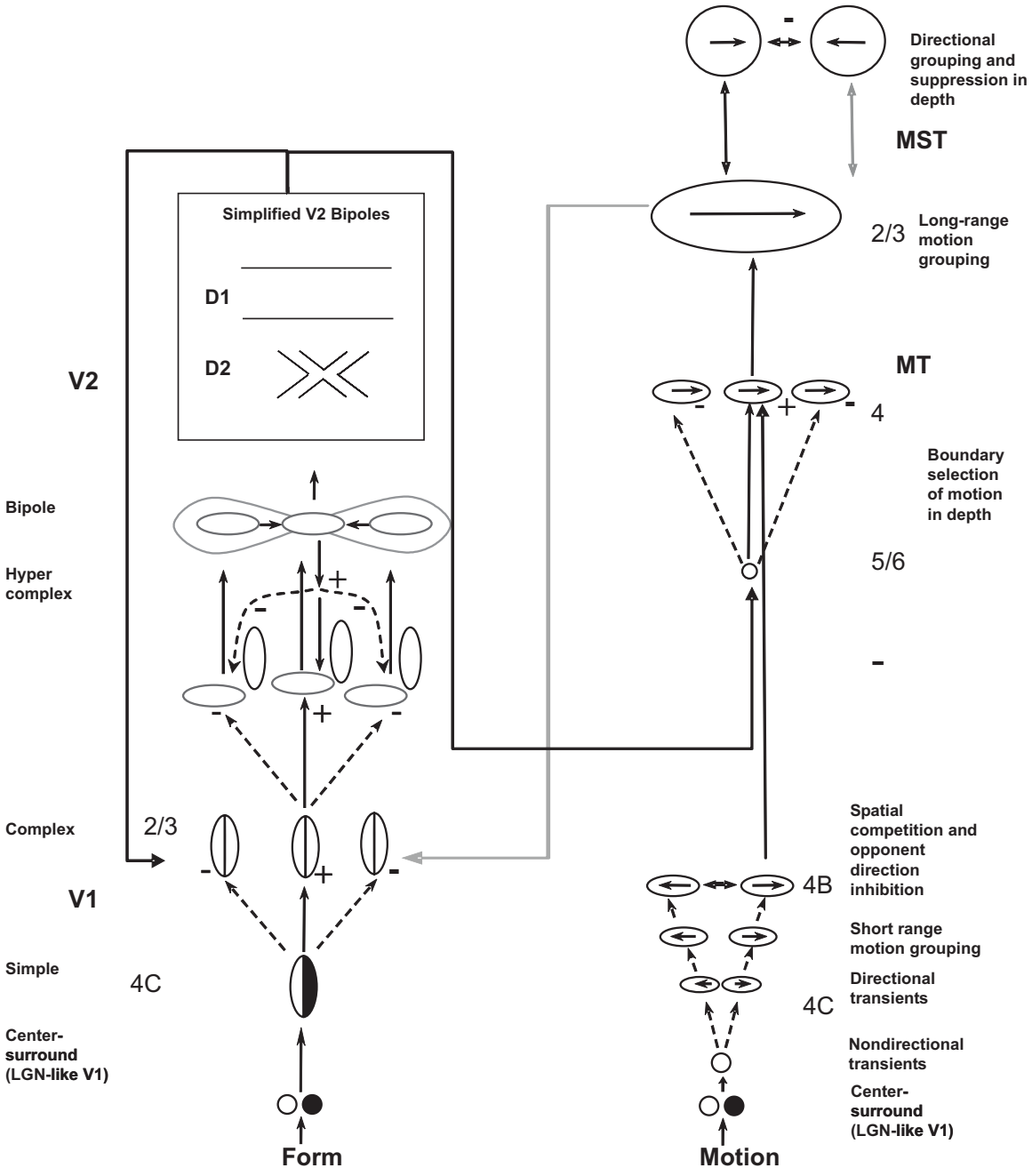

Figure 7. Laminar structure of 3D FORMOTION. See text for details.

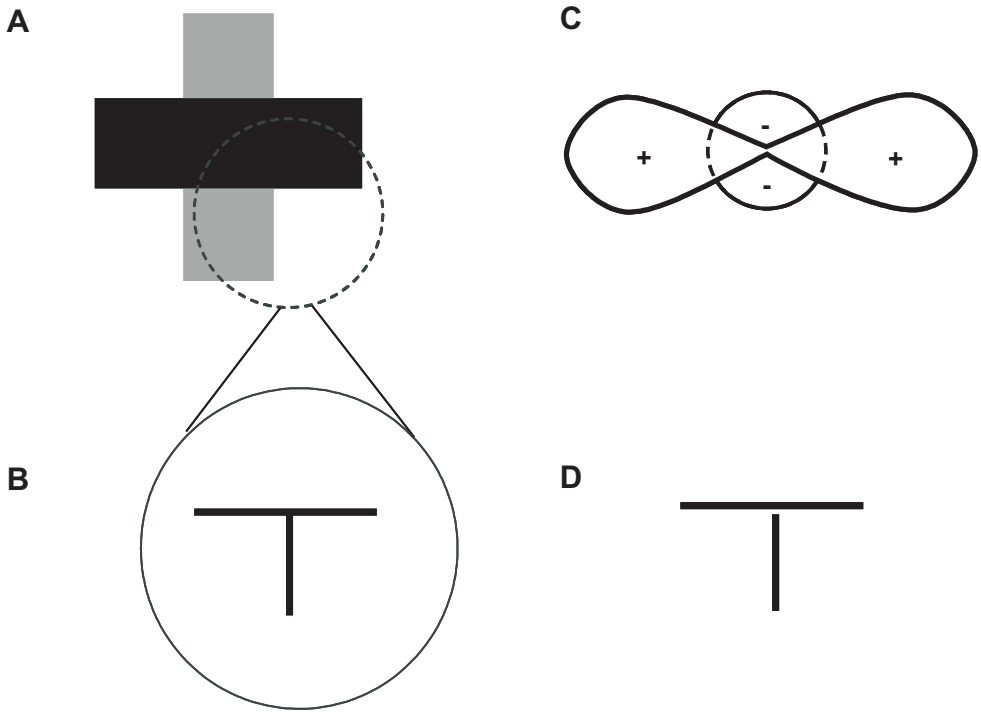

Figure 8. (A) In this 2D picture, a dark horizontal bar is perceived to be in front of a gray vertical bar. (B) The local geometry of edges in the indicated area forms a T-junction. (C) In the form stream, the "bipole" combination of long-range cooperation (indicated by the figure 8) and short-range inhibition among nearby oriented units tuned to a variety of orientations (indicated by the circle) acts at the T-junction. Only the horizontal unit is shown. (D) The result of the cooperative-competitive dynamics in (C) is that the favored collinear structure of the horizontal edge wins at the top of the T, and a small "end gap" is created at the top of the stem of the T. Due to the way in which this boundary interacts with the surface formation stream, the top of the T is assigned to the Near depth, while the vertical segment is assigned to the Far depth.

Motion Modulation of Figure-Ground Separation. Form cues are not always available to initiate figure-ground separation. Motion cues can initiate figure-ground separation even when form cues are not available. One such route in the model is via feedback projections from MT to V1 (Figures 6 and 7; Equation A28), which have been reported both anatomically and electrophysiologically (Bullier, 2001; Jones, Grieve, Wang and Sillito, 2001; Movshon and Newsome, 1996) that uses attentional biasing within MT/MST (Treue and Maunsell, 1999). How this happens is nicely illustrated by the chopsticks display in Figure 3B. Focusing spatial attention at one end of a chopstick can enhance that chopstick's direction of motion within the MT/MST complex at a given depth. Enhanced MT-to-V1 feedback can selectively strengthen the boundary signals of one chopstick in Figure 3B enough to trigger its boundary completion and figure-ground separation via V1-to-V2 interactions, even when the enhanced motion signals from this chopstick may be the only cue for depth separation in the form system. In this way, the two overlapping bars of a chopsticks display can induce separate boundaries in depth that, by closing the V2-to-MT loop, can support depth-selective motions by the chopsticks in opposite directions (Bradley, Chang and Andersen, 1998; Grossberg et al., 2001).

The Motion Processing System

The motion processing part of the model consists of six stages that represent cell dynamics homologous to LGN, V1, MT, and MST (Figure 7, right). These stages are mathematically defined in Appendix A.

Level 1: Input from LGN. A precursor of the present model (Grossberg et al., 2001) used FACADE output from V2 as the input to the Motion system. In the 3D FORMOTION model, the boundary input is not depth-specific. Rather, the 2-cell wide boundary input models the signals that come from Retina and LGN, which are lumped into a single processing stage for simplicity, into V1 (Xu, Bonds and Casagrande, 2002). This boundary is represented in both ON and OFF channels. After V1 motion processing, described below, the motion signal then goes on to MT and MST. The 3D figure-ground separated boundary inputs in the current model come from V2 to MT and select bottom-up motion inputs from V1 in a depth-selective way. This biologically more realistic input scheme proposes how the visual system separates the occluder boundaries from the moving boundaries into different depth planes, even though the inputs themselves occur within the same depth plane. The present model proposes how a combination of habituative (Equations A4—A6) and depth selection (Equation A14) mechanisms accomplishes the required depth segregation of motion signals.

These mechanisms are proposed to also play several other roles in motion processing. In particular, habituative mechanisms are part of the preprocessing whereby motion cues trigger the activation of transient cells; see below. Because the occluder boundaries are static, at least relative to the continuously moving chopsticks, their signals become much weaker over time. As a result, when the chopsticks move along the fixed locations of static occluders (Figure 3A), they generate much weaker motion signals than the same chopsticks moving without occluders (Figure 3B). This habituative property helps to explain why visible occluders generate weaker motion signals at all depth planes. It does not, however, separate intrinsic from extrinsic boundaries, and do so in depth. The motion selection mechanism does this by using depth-separated occluder and occluding boundary signals from V2 to MT. As noted above, after the BCS completes contours in corresponding depths (Equations A38 and A43), these signals are approximated by 1-pixel wide, depth-separated boundaries. The model shows how these boundaries can capture only the appropriate motion signals onto their respective depth planes in MT (see Figure 12 below).

3D FORMOTION uses both ON and OFF input cells. For example, when a bright chopstick moves to the right on a dark background (Figure 3, polarities are reversed for illustration purposes), ON cells respond to its leading edge, but OFF cells respond to its trailing edge. Likewise, when the chopstick reverses direction and starts to move to the left, its leading edge now activates ON cells and its trailing edge OFF cells. By differentially activating ON and OFF cells in different parts of this motion cycle, these cells have more time to recover from habituation, so that the system remains more sensitive to repetitive motion signals. Model ON and OFF responses are thus relevant to the role played by habituative mechanisms in generating transient cell responses and in weakening the boundaries of occluders.

Level 2: Transient cells. The second stage of the motion processing system (Figures 6 and 7) consists of non-directional transient cells, inhibitory directional interneurons, and directional transient cells. The non-directional transient cells respond briefly to a change in the image luminance, irrespective of the direction of movement (Equations A4—A6). Such cells respond well to moving boundaries and poorly to the static occluder because of the habituation, or adaptation that creates the transient response. The type of adaptation that leads to these

transient cell responses is known to occur at several stages in the visual system, ranging from retinal Y cells (Enroth-Cuggell and Robson, 1966; Hochstein and Shapley, 1976a, 1976b) to cells in V1 (Abbott, Sen, Varela and Nelson, 1997; Carandini and Ferster, 1997; Chance, Nelson and Abbott, 1998; Varela, Sen, Gibson, Fost, Abbott and Nelson, 1997) and beyond.

The non-directional transient cells send signals to inhibitory directional interneurons and directional transient cells, and the inhibitory interneurons interact with each other and with the directional transient cells (Equations A7 and A8). The directional inhibitory interneuronal interaction enables the directional transient cells to realize directional selectivity at a wide range of speeds (Grossberg, Mingolla, and Viswanathan, 2001). This predicted interaction is consistent with retinal data concerning how bipolar cells interact with inhibitory starburst amacrine cells and direction-selective ganglion cells, and how starburst cells interact with each other and with ganglion cells (Fried, Münch, and Werblin, 2002). The possible role of starburst cell inhibitory interneurons in ensuring directional selectivity at a wide range of speeds has not yet been tested.

A directionally selective neuron fires vigorously when a stimulus is moved through its receptive field in one direction (called the preferred direction), while motion in the reverse direction (called the null direction) evokes little response (Barlow and Levick, 1965). Mechanisms of direction selectivity include asymmetric inhibition along the preferred cell direction, notably an inhibitory veto of null-direction signals (Equations A7 and A8), as in Grossberg et al. (2001).

As noted above, after the transient cells adapt in response to a static boundary, then boundary segments that belong to a static occluder (extrinsic terminators, Figure 3A) produce weaker signals than those that belong to a continuously moving object. In the invisible occluder display (Figure 3B), the horizontal motion signals at the chopstick ends will be strong, and thus influence the final outcome.

Level 3: Short-range filter. A key step in solving the aperture problem is to strengthen unambiguous feature tracking signals relative to ambiguous motion signals. Feature tracking signals are often generated by a relatively small number of moving features in a scene, yet can have a very large effect on motion perception. One process that strengthens feature tracking signals relative to ambiguous aperture signals is the short-range spatial filter (Figure 7). Cells in this filter *accumulate evidence* from directional transient cells of similar directional preference within a spatially anisotropic region that is oriented along the preferred direction of the cell. This computation selectively strengthens the responses of short-range filter cells to feature-tracking signals at unoccluded line endings, object corners, and other scenic features (Equation A9). The use of a short-range spatial filter followed by competition at Level 4 eliminates the need for an explicit solution of the *feature correspondence problem* that various other models posit and attempt to solve (Reichardt, 1961; van Santen and Sperling, 1985).

Level 4: Spatial competition and opponent direction competition. Two kinds of competition further enhance the relative advantage of feature tracking signals (Figures 6 and 7, Equation A11). These competing cells are proposed to occur in layer 4B of V1 (Figure 7; bottom -right). Spatial competition among cells of the same spatial scale that prefer the same motion direction boosts the amplitude of feature-tracking signals relative to those of ambiguous signals. Feature tracking signals are contrast-enhanced by such competition because they are often found at motion discontinuities, and thus get less inhibition than ambiguous motion signals that lie within an object's interior. Opponent-direction competition also occurs at this processing stage, with properties similar to the V1 cells described by Rust, Majaj, Simoncelli and Movshon (2002) both in exhibiting an opponent direction mechanism, and in having the correct spatial scale for

such interactions.

The activity pattern at this model stage is consistent with data of Pack, Gartland, and Born (2004). First, in their experiments, V1 cells demonstrate an apparent suppression of responses to motion along visible occluders. A similar suppression occurs in the model due to the adaptation of transient inputs to static boundaries. Second, cells in the middle of a grating (influenced only by ambiguous signals) respond more weakly than cells at the edge of the grating (influenced by intrinsic terminators). This effect is explained in the model by spatial competition between motion signals. This process performs divisive normalization and endstopping, which together serve to amplify the strength of directionally unambiguous feature tracking signals at line ends relative to the strength of aperture-ambiguous signals along line interiors.

Level 5: Long-range filter and formotion selection. Motion signals from model layer 4B of V1 input to model area MT. Area MT also receives a projection from V2 (Anderson and Martin, 2002; Rockland, 1995) that carries depth-specific figure-ground-separated boundary signals. These V2 form boundaries select the motion signals (*formotion selection)* by selectively assigning to different depths the motion signals coming into MT from layer 4B of V1 (Equation A14). When the dynamically formed V2 boundary signals satisfy an appropriate criterion (Equations A38 and A43), they are projected to MT as idealized depth-separated boundaries. This approximation eliminates the need to do a complete FACADE model simulation.

Formotion selection, or selection of motion signals in depth by corresponding boundaries, is proposed to occur via a narrow excitatory center, broad inhibitory surround projection from V2 to layer 4 of MT. For example, in response to the chopsticks display with visible occluders (Figure 3A), the formotion selection mechanism for depth D1 selects motion signals at its positions in D1, which lie along the visible occluder boundaries, and suppresses motion signals at other locations in depth D1. The resulting activation in D1 will be weak, due to the habituated bottom-up input from V1 along the selected occluder boundary positions (see Figure 14A in the Results section). The V2 boundary signals that correspond to the moving boundaries select strong motion signals at depth D2 (see Figure 14B in the Results section).

A similar type of inter-stream gating signal is proposed to play a key role in explaining challenging data about stereopsis, 3D surface perception, and figure-ground separation (Cao and Grossberg, 2005; Fang and Grossberg, 2004; Grossberg, 1994, 1997; Grossberg and Yazdanbakhsh, 2005). This gating signal is proposed to operate within the form system, namely from the thin stripes to the pale stripes of V2, and allows 3D surface feedback to modulate the strength of 3D boundaries that control visible 3D form percepts. Thus it seems that several different types of gating occur across the parallel visual processing streams at the V2 and MT processing levels.

The boundary-gated signals from layer 4 of MT are proposed to input to the upper layers of MT (Figure 7, top-right), where they activate directionally-selective, spatially *anisotropic* filters via long-range horizontal connections (Equation A16). In this *long-range filter,* motion signals coding the same directional preference are pooled from object contours with multiple orientations and opposite contrast polarities. This pooling process creates a true directional cell response (Chey et al., 1997; Grossberg et al., 2001; Grossberg and Rudd, 1989, 1992). Earlier versions of the long-range filter used a spatially isotropic kernel, for simplicity. In order to explain the types of data analyzed in this paper, we propose that the long-range filter accumulates evidence of a given motion direction using a kernel that is elongated in the direction of that motion, much as in the case of the short-range filter. This hypothesis is consistent with data showing that approximately 30 % of the cells in MT show a preferred direction of motion

that is aligned with the main axis of their receptive fields (Xiao, Raiguel, Marcar and Orban, 1997).

The predicted long-range filter cells in layer 2/3 of MT are proposed to play a role in binding together 3D directional information that is homologous to the orientationally selective, coaxial and collinear accumulation of evidence within layer 2/3 of the pale stripes of cortical area V2 for the purpose of 3D perceptual grouping of form (Grossberg 1999; Grossberg and Raizada, 2000). This anisotropic long-range motion filter allows motion signals to be selectively integrated across occluders with variable degrees of success in response to the various shapes in the Lorenceau-Alais displays of Figure 4.

Level 6: Directional grouping. The model processing stages up to now do not fully solve the aperture problem. Although they can amplify feature tracking signals and assign motion signals to the correct depths, they cannot yet explain how feature tracking signals can propagate across space to select consistent motion directions from ambiguous motion directions, without distorting their speed estimates, and at the same time suppress inconsistent motion directions. They also cannot explain how motion integration can compute a vector average of ambiguous motion signals across space to determine the perceived motion direction when feature tracking signals are not present at that depth. The final stage of the model accomplishes this goal by using a motion grouping network (Equations A16 and A21), interpreted to occur in ventral MST (MSTv). We predict that this motion grouping network determines the coherent motion direction of discrete moving objects.

The motion grouping network works as follows: Cells that code the same direction in MT — and also perhaps similar directions, but this possibility is not explored herein — send convergent inputs to cells in MSTv via the motion grouping network. Within MSTv, directional competition at each position determines a winning motion direction. This winning directional cell then feeds back to its source cells in MT. This feedback supports the activity of MT cells that code the winning direction, while suppressing the activities of cells that code all other directions. This motion grouping network enables feature tracking signals to select similar directions at nearby ambiguous motion positions, while suppressing other directions there.

On the next cycle of the feedback process, these newly unambiguous motion directions select consistent MSTv grouping cells at positions near them. The grouping process propagates across space as the feedback signals cycle through time between MT and MSTv. Chey et al. (1997) and Grossberg et al. (2001) first used this process to simulate data showing how the present model solves the aperture problem, and Pack and Born (2001) have recently provided supportive data, by showing that the response of MT cells to the motion of the interiors of extended lines is over time dynamically modulated away from the local direction that is perpendicular to the contour and towards the direction of line terminator motion.

It is worth noting that both the V2-to-MT and the MSTv-to-MT signals carry out selection processes using modulatory on-center, off-surround interactions. The V2-to-MT signals select motions signals at the locations and depth of a moving boundary. The MST-to-MT signals select motion signals in the direction and depth of a motion grouping. Such a modulatory on-center, off-surround network was predicted by Adaptive Resonance Theory to carry out attentive selection processes in a manner that enables fast and stable learning of appropriate features to occur. See Raizada and Grossberg (2003) for a review of behavioral and neurobiological data that support this prediction in several brain systems. Direct experiments to test it in the above cases still remain to be done.

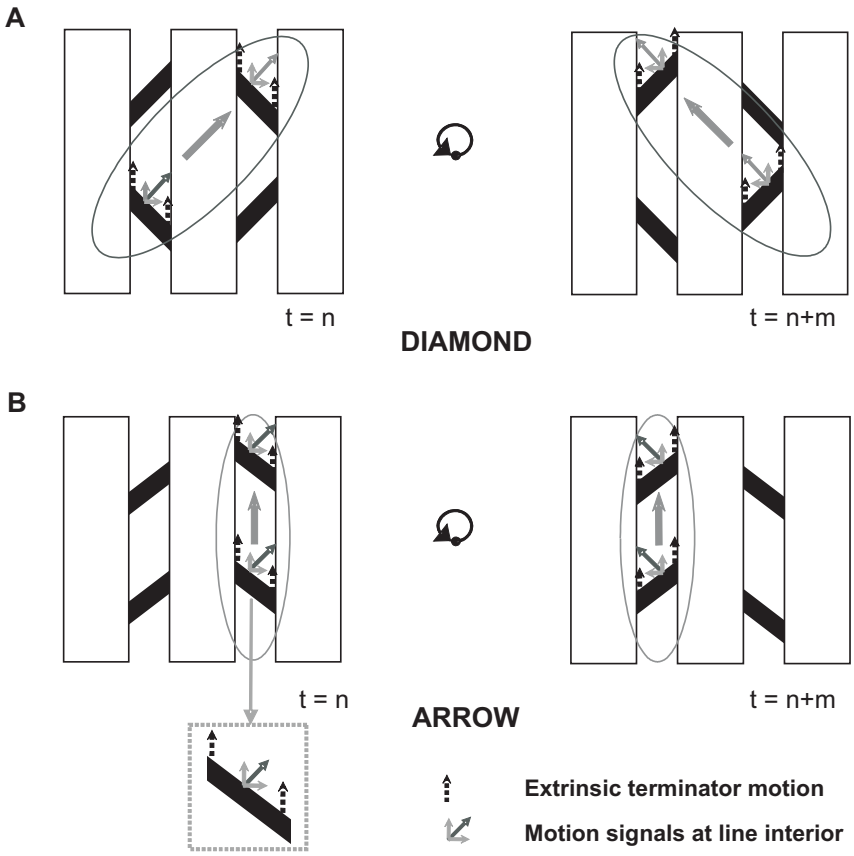

Figure 9. Motion signals in Diamond (A) and Arrow (B) displays with visible occluders. Ellipses represent receptive fields of long-range motion grouping MT cells (with direction preference indicated by the large gray arrow), that are activated the most by the given combination of motion signals. Counterclockwise motion direction is indicated by the circular arrow in the middle. At time *t=n,* both diamond and arrow centers move along the bottom-right quadrant of the circular trajectory, and global motion of the input stimulus is up-right (45º). At time *t=n+m,* global motion of the stimulus is up-left (135º). The motion grouping is consistent with the globally perceived motion only in the diamond display. See text for details.

Analysis and Simulation of Psychophysical Experiments

This section is devoted to a detailed analysis and simulations of three important kinds of psychophysical displays: shapes moving behind occluders (Lorenceau-Alais, 2001), chopsticks (after Anstis, 1990) and rotating ellipses (Weiss and Adelson, 2000).

Movement behind occluders. Lorenceau and Alais (2001) created displays in which circular-parallel motion was visible through the two vertically oriented apertures, but the corners of the shapes remained hidden (Figure 4). See http://cns.bu.edu/~juliaber/formotion.html. Therefore, observers had to rely on motion integration across space to determine motion direction. The success of the motion integration process depended on the type of shape and on the contrast of the occluders. The diamond displays resulted in a higher percentage of correct responses than the cross and arrow displays, and displays with visible occluders were easier than

those with invisible ones. For example, a diamond (Figure 4A) rotating behind visible occluders created a percept of a *single* rotating shape. In contrast, a rotating arrow (Figures 4C and 4F) produced a percept of two *disconnected* shapes separately moving in their respective apertures. This disconnection was strong even in the case of visible occluders (Figure 4C) and more pronounced in the case of invisible occluders (Figure 4F).

Schematic representations of the motion grouping signals generated by the displays of a diamond and an arrow with visible occluders are shown in Figure 9. Both shapes undergo a counterclockwise motion (as denoted by a circular arrow in the middle). At the corresponding time points (for example, Figures 9A and 9B, $t = n$), each display has a combination of the same set of local motion signals. Perceptual dissimilarities are caused by the difference in relative positioning of those motion signals through time. Both the diamond and the arrow are visible through the apertures as four linear boundary segments. Each segment produces two types of motion signals: ambiguous signals (due to the aperture problem) from line interiors and unambiguous signals from terminators (Figure 9 inset). For the visible occluder cases, the terminator signals are extrinsic and weak. Ambiguous motion signals of the same direction from parallel segments can then combine across space using the model's anisotropic motion grouping filters to produce the perceived object motion.

For example, in the diamond display in Figure 9A, two line segments with synchronous motion in a given direction are located in different apertures. The large anisotropic motion grouping cells that prefer this motion direction can thus integrate the diagonal motion signals across the apertures. At time $t = n$, when the diamond center traverses a bottom-right trajectory quadrant, two segments moving simultaneously in the up-right (45º) direction activate the diagonal motion cells, while only one segment activates vertical or horizontal ones. The MT-MST motion grouping network therefore prefers the diagonal signals from the line interiors to the weaker vertical or horizontal groupings. Cells activated the most would be those over the center of the rotating shape. First, the cells with a 45º (up-right) direction preference will be activated ($t = n$), then 135º (up-left) cells ($t = n + m$), 225º, 315º, and then back to the beginning of the cycle. Simulation results are shown in Figure 10. This sequence of motion signals is consistent with the circular-parallel motion in a counter-clockwise direction, leading to a coherent percept of a rotating diamond.

For the arrow display in Figure 9B, vertical components of the ambiguous signals from the line interiors and vertical extrinsic signals from the line ends activate vertically oriented anisotropic long-range filter cells. Diagonal ambiguous motion signals from neighboring parallel shape segments can only weakly group together within one aperture, and so lose the directional competition that determines the winning direction. As a result, a vertical (upward) direction of motion will accumulate in the right aperture ($t = n$) when the arrow center traverses the bottom-right trajectory quadrant, but at a later time ($t = n + m$), top-right trajectory quadrant, this vertical direction will develop in the left aperture The result is a seesaw up-and-down translational motion that is inconsistent with rotation. Such out-of-phase timing of motion signals will prevent motion integration across the two apertures. Another way of saying this is that asynchronous motions of similar directions produce a segmentation signal, thus preventing a percept of a single rotating object.

Analysis of motion signals in the invisible occluder displays (Figures 4D-4F) is similar to the analysis above. Because line terminators are intrinsic, they will produce stronger vertical signals and aid the vertical motion grouping. Simulations of motion segmentation for the case of arrow with invisible occluders (Figure 4F) are shown in Figure 11.

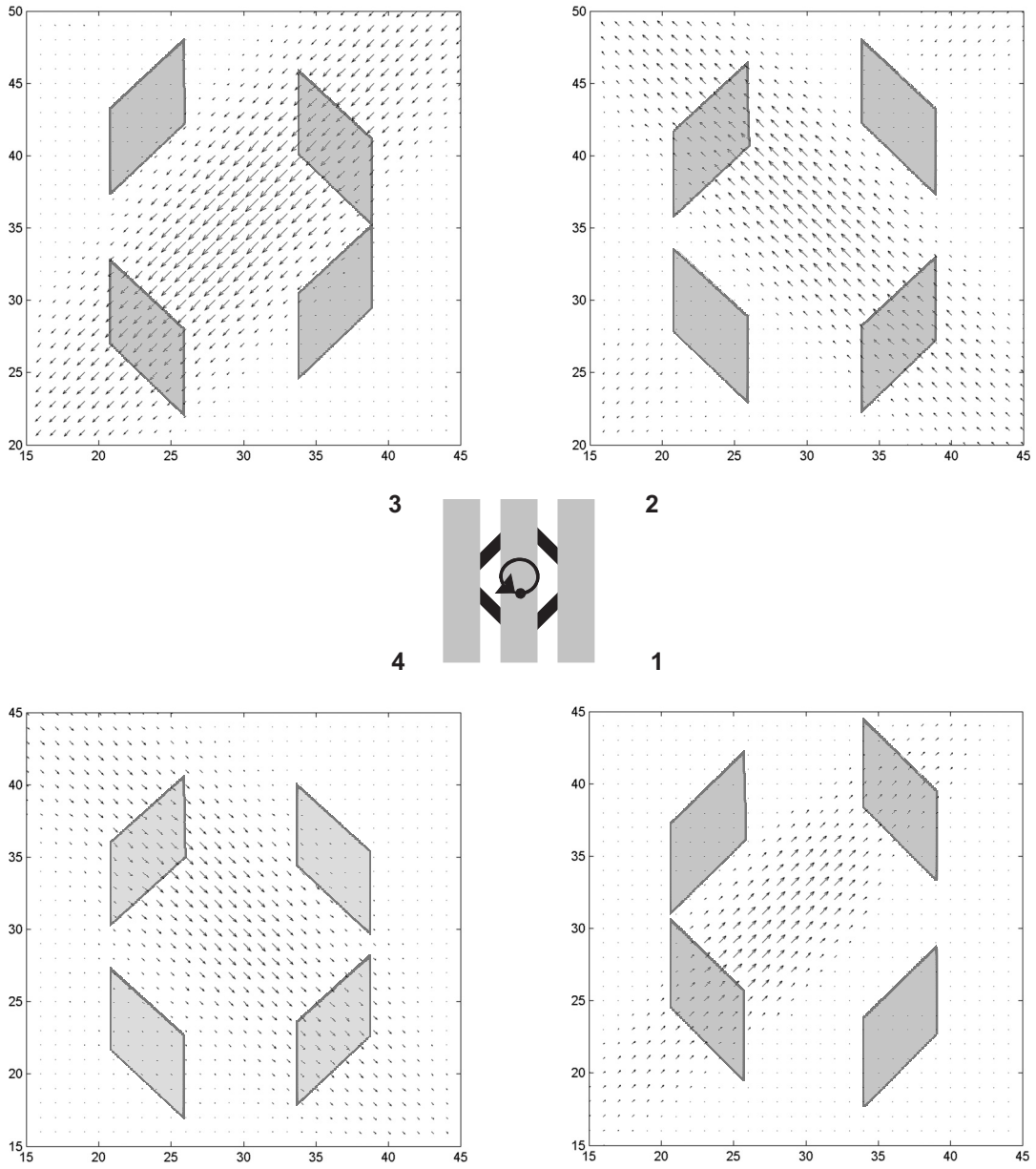

Figure 10. Simulation of motion signals in Diamond display with visible occluders. MT output (Motion Level 5, Equation A16) in depth D2 for a sequence of four frames (1,2,3,4) in four quadrants (bottom right, top-right, top-left and bottom-left) of the circular trajectory. This sequence of motion signals is consistent with a circular motion of a single shape. Direction and length of individual arrows represent the direction and strength of MT cell activation at each point.

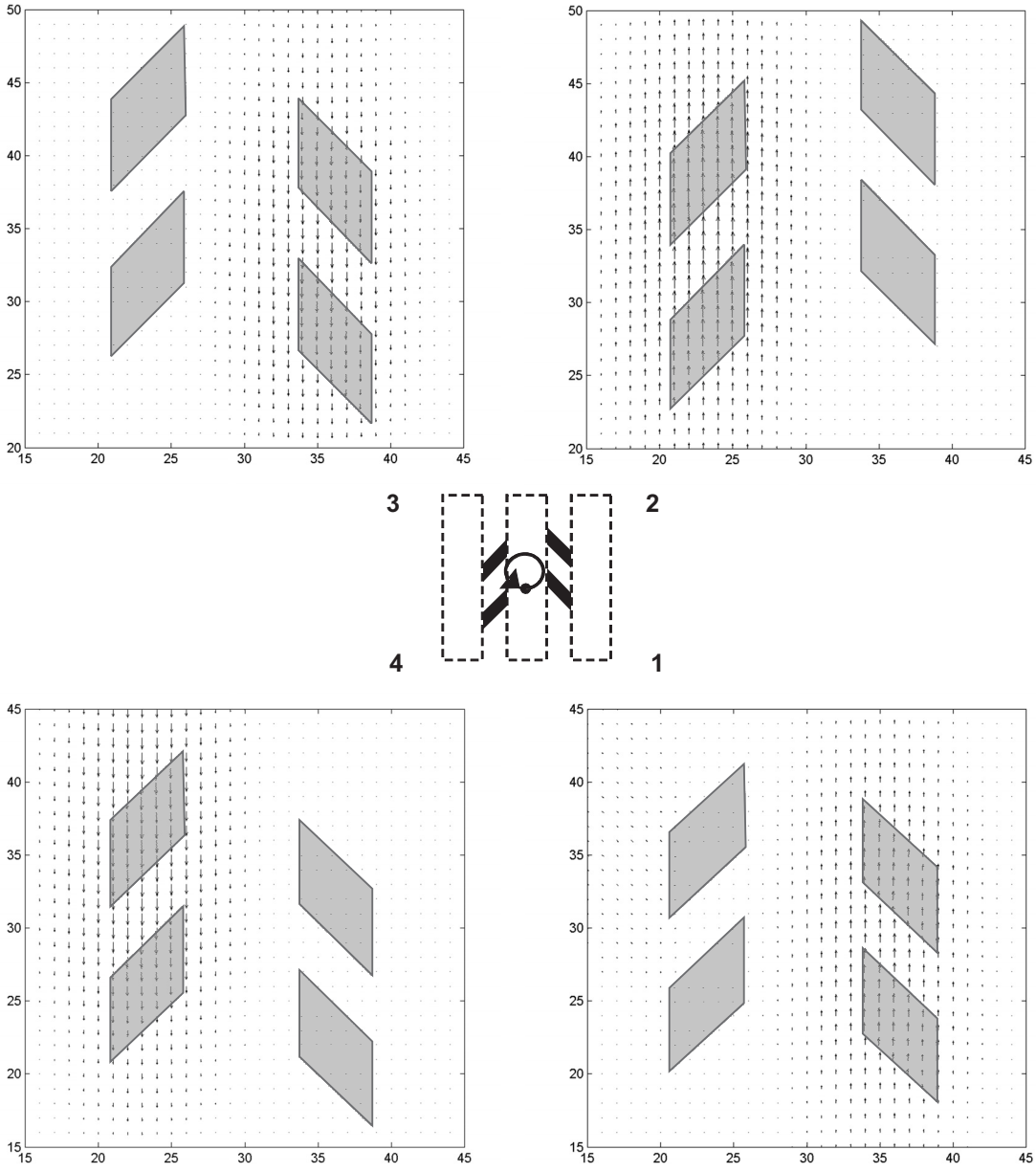

Figure 11. Simulation of motion signals in Arrow display with invisible occluders. MT output (Motion Level 5, Equation (A16)) in depth D1 for a sequence of four frames (1,2,3,4) in four quadrants (bottom right, top-right, top-left and bottom-left) of the circular trajectory. This sequence is consistent with a translational motion of two separate shapes. Direction and length of individual arrows represent the direction and strength of MT cell activation at each point.

An intermediate image configuration, such as the diamond with invisible occluders in Figure 4D, creates strong vertical feature tracking signals within each aperture that can better compete with the strong diagonal ambiguous motion grouping across apertures. The percept is thus determined by competition between two motion directions and results in a larger number of "incorrect answers" than does the percept in the visible occluder case of Figure 4A. In the case of an arrow with visible occluders in Figure 4C, the vertical signals will be weak because they are extrinsic, whereas in Figure 4F they are strong because they are intrinsic. Thus, translation will overwhelm rotation less in Figure 4C than in Figure 4F, and the number of correct responses about arrow rotation will be higher there. All of these model properties are consistent with the data of Lorenceau and Alais (2001).

Chopsticks with visible and invisible occluders. Two configurations of the chopsticks display, with visible and invisible occluders (Figure 3A and 3B), were simulated. See http://cns.bu.edu/~juliaber/formotion.html. In the case of visible occluders, chopsticks are perceived moving coherently in a vertical direction. In the case of invisible occluders, the percept is of two horizontally moving objects, one moving in front of the other. These two displays differ only at the chopsticks' ends. The difference in motion percept here can be explained by the difference in the relative strength of unambiguous feature-tracking motion signals of the intersection and either strong (intrinsic) or weak (extrinsic) motion of the chopsticks' ends. Aperture-ambiguous motion signals at the line interiors do not play a significant role in this percept.

Independent of the visibility of occluders, in the *static* image, the two chopsticks are perceived as one X-shaped pattern. However, in the moving image, chopsticks with invisible occluders separate in depth and are perceived as sliding one above another. Simulations of the chopstick display in the invisible occluder case are shown in Figures 12 and 13. Figure 12 shows how, in the motion system, opposite direction signals from two chopsticks separate in depth. The sequence of motion computations leading to this percept starts with strong horizontal motion direction signals from the intrinsic terminators at the chopsticks' ends. These feature-tracking signals are amplified by anisotropic short-range motion filters of V1 that accumulate evidence in a given motion direction as the chopstick moves along, and are integrated by the long-range filters of MT. Attentional priming biases motion signals at one chopstick end (top-left) in the near depth. Competition within the MT-MST circuit includes asymmetric inhibition from the near depth (D1) to the far depth (D2) ("asymmetry between near and far"). This interaction results in the primed motion direction winning in D1 and another motion direction winning in D2. Attentionally biased competition in the motion stream is similar to the proposed effect of attention in the form stream (Carpenter and Grossberg, 1991; Grossberg, 1980; Reynolds, Chellazi and Desimone, 1999).

Initially, the bipole cells of orthogonal diagonal orientation preferences in the V2 form system compete with each other, but are unable to complete over the gap formed by the chopsticks' intersection (Figure 13A). The bias that allows one chopstick to win the competition can be provided by an attentional input to the form system, by an attentional input to the motion system that is fed back from the motion system to the form system, or by introducing some inequality in the chopsticks' physical properties (e.g., by making one thicker).

In the current simulations, depth-selective attentive feedback from MT modulates complex cells of the corresponding depths in V1. This feedback equals the sum of the motion signals at a given depth, and is not orientation-selective or direction-selective. Motion signals in MT are spatially restricted to one chopstick in each depth and, through the feedback, enhance

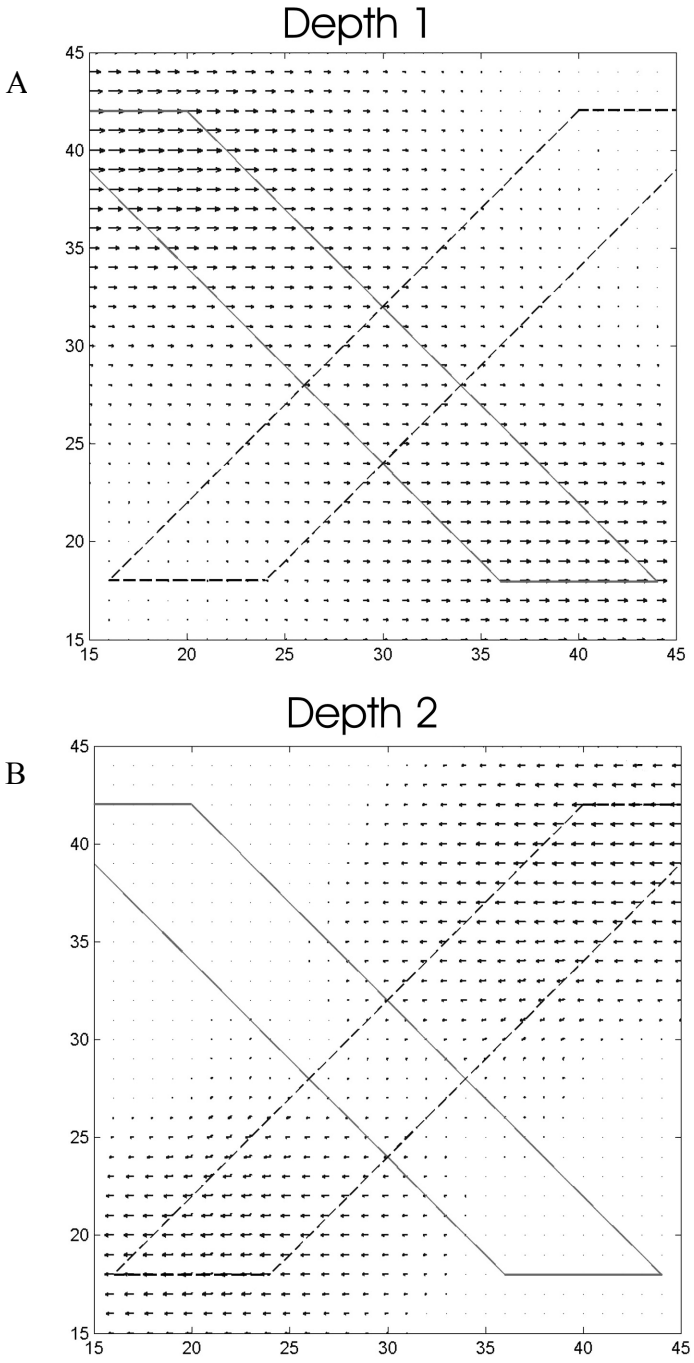

Figure 12. Motion computation in MT (Motion Level 5, Equation (A16)) for chopsticks with invisible occluders. Rightward motion of one chopstick is represented in depth level D1 (A), and leftward motion of the second chopstick at the depth level D2 (B). Direction and length of individual arrows represent the direction and strength of MT cell activation at each point.

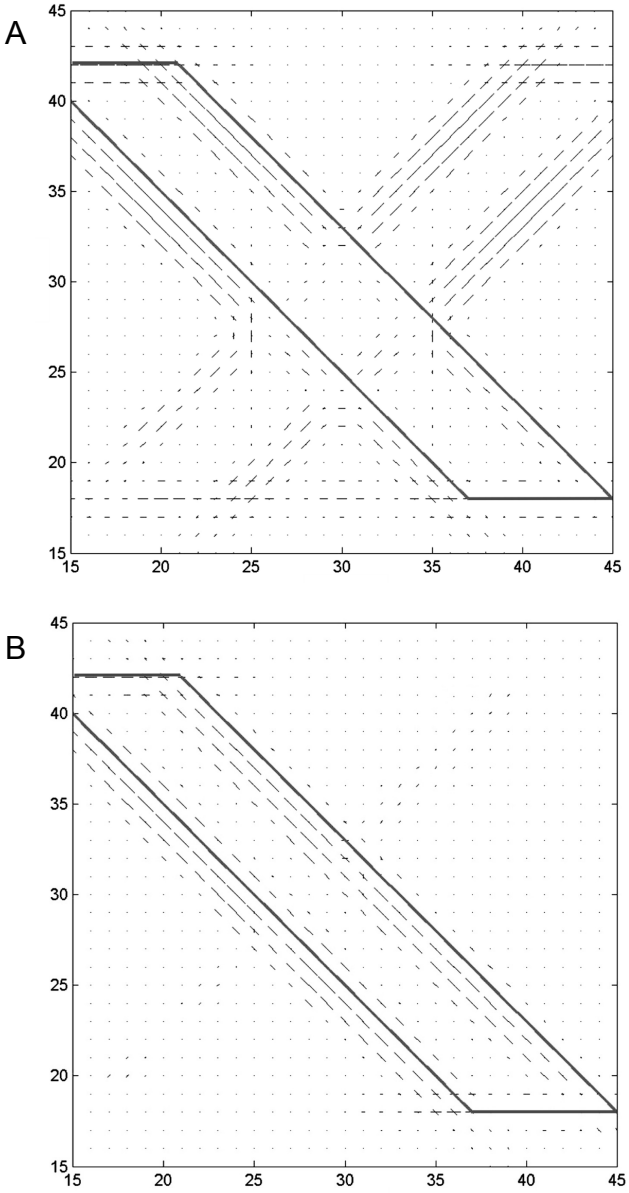

Figure 13. Boundary computation (bipole output, Form Level 5, Equation (A38)) for chopsticks with invisible occluders. Spatial scale 1 is shown. (A) Initially, there is no separation of boundaries between occluder and occluded objects. (B) Bias from motion system can strengthen boundary inputs in a topographic manner, and allow one chopstick boundary to win and complete in D1. Orientation and length of short individual lines represent the orientation and strength of bipole cell activations at each position. The rectangular outline represents the location of the left bar in the chopsticks display.

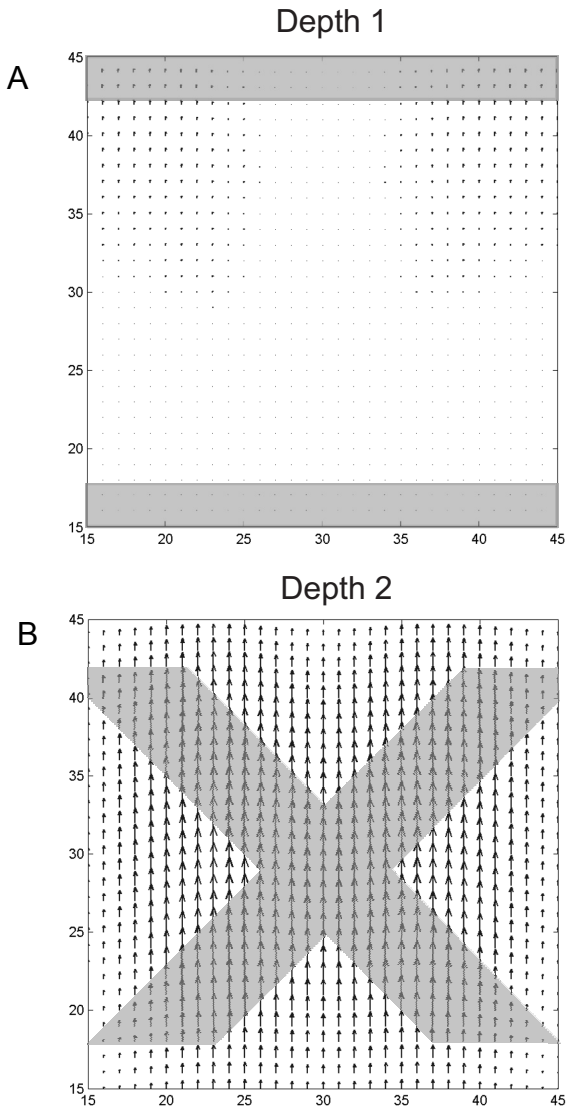

Figure 14. Motion computations in MT (Motion Level 5, Equation (A16)) for chopsticks display with visible occluders. (A) Boundaries in the near depth (D1) select only a weak motion signal, and suppress a signal in the middle of the display. (B) Coherent motion signal is computed in the farther depth D2. Direction and length of individual arrows represent the direction and strength of MT cells activation at each point.

boundary signals for this chopstick more than for the other. Due to this motion bias, boundaries of the corresponding chopstick complete in the near depth, D1 (Figure 13B), thus pushing the second chopstick boundary in the further depth via mechanisms that are simulated in the full FACADE model (e.g., Grossberg and Yazdanbakhsh, 2005; Kelly and Grossberg, 2000). Here, we use an algorithmic separation of boundaries in depth as soon as the bipole activation

(Equation (A38)) of the attended chopstick completes over the ambiguous gap where the two chopsticks cross (see Appendix, Equation (A43)).

In the case of visible occluders, the chopsticks' ends are extrinsic terminators and do not create strong motion signals, but the vertical motion of the chopstick intersection is unambiguous and strong. The result of motion integration and competition is a coherent, vertical motion signal at the far depth, D2 (Figure 14B). This signal does not provide a segmentation bias in feedback from MT to the form system. The form system output at the far depth, D2, is an outline of an "X" shape (Figure 7, V2, top-left) moving up and down, and none of the competing boundaries is able to win. The form system output at the near depth, D1, consists of two static horizontal boundaries of the occluders (Figure 7, V2, top-left). The model predicts that these depth-separated boundaries in V2 select motion signals in the corresponding depth representations of MT via V2-to-MT projections with excitatory centers and inhibitory surrounds; that is, via the modulatory on-center, off-surround network. Bottom-up motion signals along the horizontal occluder boundaries consist mainly of the motion of extrinsic terminators, and are weakened by adaptation at the input layers of V1 (transient cells in Figure 6). Furthermore, surround inhibition produced by the same boundaries suppresses motion signals from interior parts of the display. This combination of narrow excitatory projections from V2 to MT with wide inhibitory surrounds results in no significant motion signal in the MT representation of the near depth, D1 (Figure 14A). On the other hand, selection by "X"-shaped boundaries in D2 picks up a strong bottom-up signal from the chopsticks' intersection and the selected vertical ambiguous signals from line interiors, resulting in a global vertical motion percept in the far depth, D2.

Gelatinous ellipses. The perception of rigidity of rotating ellipses depends on their shape (Figure 5). The 3D FORMOTION model suggests that the processes determining rigidity of the boundary are similar to those determining the percept of coherent *vs*. incoherent motion, as well as the percept of a single object *vs.* assignment of neighboring boundaries to different objects, possibly at different depths. In the non-rigid case (thick ellipse), analysis of local motion signals shows that local motion signals perpendicular to the ellipse boundary may prevail. As in the case of incoherent Lorenceau-Alais displays (arrow), each segment of the ellipse boundary moves in the manner inconsistent with a single (object) motion in the display (Figure 15A).

In the rigid ellipse case (thin ellipse) the dominant motion signal is consistent with a single object rotation that is tangential to the boundary at the points of the highest curvature (Figure 15B). The resulting motion percept in the ellipse displays is determined by the competition among ambiguous local signals integrated through large MT receptive fields. This hypothesis is supported by the "satellite effect" (Weiss and Adelson, 2000): dots moving outside of the ellipse can bias the perception of rigidity. If dots, which provide unambiguous motion signals, move along circular trajectories, then the ellipse, even a thick one, is perceived as rigidly rotating (Figure 16A). If dots oscillate in the direction orthogonal to the contour, the ellipse, even a thin one, is perceived as deforming (Figure 16B).

Weiss and Adelson (2000) reported that the capture of an ambiguous ellipse motion by unambiguously moving satellites happens even if both lie in different depth planes (as defined by disparity). Moreover, in the case of two pairs of satellites, the closer one in depth captures ambiguous ellipse motion and determines the global percept. These data can be explained by the depth-selectivity of V2 → MT projections (Bradley and Andersen, 1998). For example, the maximum capture signal will be at the depth of the satellites, and the strength of the capture signal will decrease with the difference in depth between the satellites and the ambiguous motion signals. The ambiguous motion signals that are closest to the depth of the satellites will thus be

captured more easily within their depth plane. The outcome of the competition of two sets of satellites will be determined by the one with the stronger motion signal in the ellipse depth plane. These effects are not simulated in the present article, but they are clearly implied by the 3D FORMOTION model.

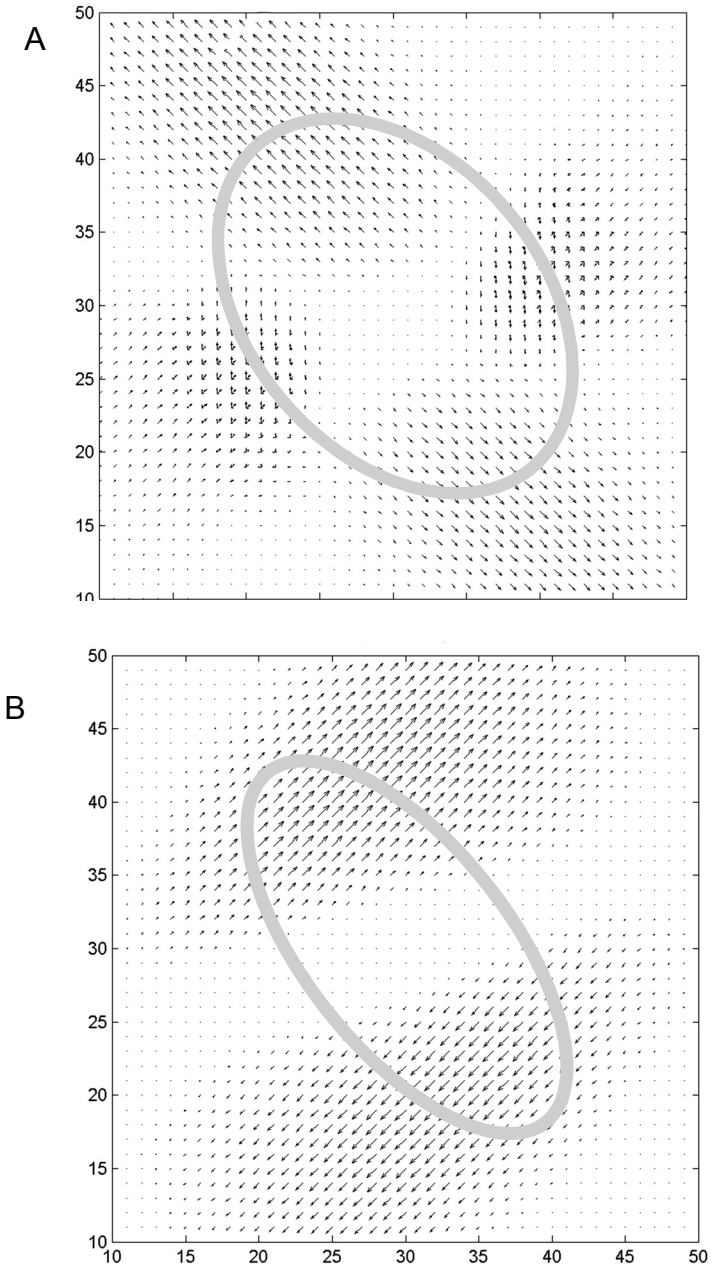

Figure 15. Motion computation in MT (Motion Level 5, Equation (A16)) in the ellipse display. (A) Thick ellipse. Motion signals are consistent with stretching of the boundary and with nonrigid percept. (B) Thin ellipse. Motion signals are consistent with rotation. Direction and length of individual arrows represent the direction and strength of MT cells activation at each point.

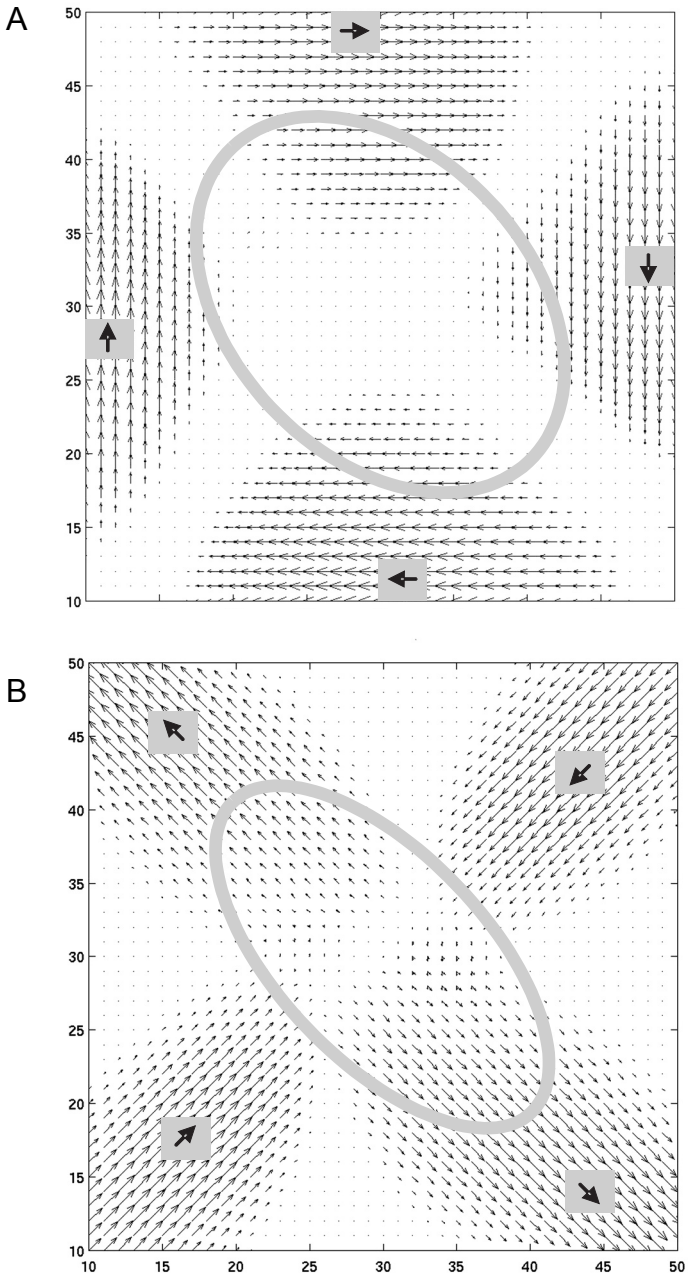

Figure 16. Motion computation in MT (Motion Level 5, Equation (A16)) in the ellipse display with satellites. Small arrows within each satellite represent direction of satellite movement (A) Thick ellipse, rotating satellites. Motion signals are consistent with rotation and with rigid percept. (B) Thin ellipse, stretching/contracting satellites. Motion signals are consistent with deformation Direction and length of individual arrows represent the direction and strength of MT cells activation at each point.

Discussion

The 3D FORMOTION model is firmly grounded in neurophysiological data. To explain psychophysical results, 3D FORMOTION predicts that a number of functional properties arise from known neural circuits of the primate motion system. Table 1 summarizes the key physiological projections and neuron properties employed by the model, alongside selected references supporting those connections or functional properties. Table 1 also lists the model's key physiological predictions that remain to be tested.

Previous models of motion integration and segmentation. A number of motion models have dealt with mechanisms of directional selectivity, motion integration and segmentation. For a review, see Grossberg et al. (2001). Few of them have addressed the issues of extrinsic *vs.* intrinsic terminators, and the effect of this dichotomy on motion processing. Lidén and Pack (1999) proposed that T-junctions, which indicate occlusion in 2D images of 3D scenes, can suppress motion signals in their vicinity. Their model does not, however, explain how occluding and occluded objects are separated in depth, or how varying the relative contrasts at X-junctions and T-junctions can cause totally different outcomes, such as perceived occlusion or transparency, as explained in Grossberg and Yazdanbaksh (2005). Wilson, Ferrera and Yo (1992) proposed that there are parallel Fourier and non-Fourier channels in motion processing. However psychophysical data do not support the existence of these pathways (Bowns, 1996; Cox and Derrington, 1994).

Authors of the three sets of data simulated in this article proposed explanations for their respective data that differ from explanations offered by the 3D FORMOTION model. For example, Lorenceau and Alais (2001) suggested that some shapes rotating behind occluders produce weak rotational motion percepts because of a "veto" imposed on motion integration. Only the "bad" shapes, those that cannot form a closed contour, would veto motion integration. Mechanisms and cortical locations of the veto process were not specified. In contrast, the 3D FORMOTION model suggests that anisotropic receptive fields integrate motion across apertures as a part of the basic process that solves the aperture problem by generating a coherent object motion percept. Some MT cells have elongated receptive fields (Xiao et al., 1997) that can be formed by long-range anisotropic projections (Schmidt, Goebel, Lowell, Singer, 1997; Sincich and Blasdel, 2001) in the upper laminae of MT (Malach, Schriman, Harel, Tootell and Malonek, 1997). The 3D FORMOTION model thus explains differences in motion percepts using known cortical mechanisms, and predicts that a correlate of coherent object motion can be found in some cells of the MT-MST grouping network.

Several prior models compute motion signals for gratings and plaids. However, none of them can explain in detail the different percepts for the chopstick illusion, which can be considered as a limiting case of a plaid consisting of just two bars: the visible occluder case produces coherent vertical motion, while the invisible occluder case results in motion separation in depth. Typically, alternative motion models concentrate on motion mechanisms and do not explain how 3D figure-ground separation mechanisms form extrinsic and intrinsic terminators, and how these terminators affect global motion computations. Grossberg et al. (2001) provided a partial explanation of how local motion signals in the ambiguous positions can be overwhelmed by the propagation of the strong feature-tracking signals from the chopsticks' ends. The 3D FORMOTION model uses the same propagation of feature-tracking signals, together with the new form-motion interactions, to more fully explain all aspects of the chopsticks illusion.

Previous models of the ellipse illusion have either accounted for the differences between rigid and nonrigid cases, but not for the effect of satellites (Hildreth, 1983), or for the effect of

Table 1

Functional projections and properties of model cell types and predictions

Connection/Functional property	**Selected references**
Functional projections	
V1 4Ca to 4B	Yabuta et al., 2001, Yabuta & Callaway, 1998
V1 to MT	Anderson et al.,1998; Rockland, 2002; Sincich & Horton 2003, Movshon & Newsome, 1996
V1 to V2	Rockland, 1992, Sincich & Horton, 2002
V2 to MT	Anderson & Martin, 2002; Rockland 2002; Shipp & Zeki 1985; DeYoe & Van Essen 1985
MT to V1 feedback	Shipp & Zeki 1989; Callaway 1998; Movshon & Newsome 1996; Hupé et al., 1998
V2 to V1 feedback	Rockland & Pandya, 1981; Kennedy & Bullier 1985
Properties	
V1 adaptation	Abbott et al.,1997; Chance et al., 1998; (rat); Carandini & Ferster, 1997, (cat)
V1(4ca) transient nondirectional cells	Livingstone & Hubel, 1984
V1 spatially offset inhibition	Livingstone, 1998; Livingstone & Conway, 2003; Murthy & Humphrey, 1999 (cat)
V2 figure-ground separation	Zhou et al., 2000; Bakin et al., 2000
MT figure-ground separation and disparity sensitivity	Bradley et al., 1998, Grunewald et al., 2002; Palanca & DeAngelis 2003
MT center- surround receptive fields	Bradley & Andersen, 1998; Born, 2000; DeAngelis & Uka, 2003
Some MT receptive fields elongated in preferred direction of motion	Xiao et al.,1997
Attentional modulation in MT	Treue & Maunsell, 1999
Predictions	
Short-range anisotropic filter in V1 (motion stream)	
Long-range anisotropic filter in MT (motion)*	
V2 to MT projection carries figure-ground completed-form-in-depth separation signal	
MT to V1 feedback carries figure-ground separation signal from motion to form stream	
MST to MT feedback helps solve aperture problem by selecting consistent motion directions	

*Although Xiao et al, 1997 found that some MT neurons have receptive fields that are elongated along the preferred direction of motion, there is no direct evidence that these neurons participate preferentially in motion grouping.

satellites but not of background motion (Grzywacz and Yuille, 1991). Multiple depth layers in combination with a smoothness constraint helped Weiss et al. (2000) to explain a rigidity percept as a function of the aspect ratio, the effect of satellites, and the effect of a background motion. That work, however, did not suggest a neural implementation. Our model suggests specific mechanisms: depth-specific boundary selection of motion, together with motion integration and segregation mechanisms, allows it to address all variations of the ellipse display.

A number of more recent models of vision employ Bayesian techniques. One that is particularly relevant to this work is that of Weiss, Simoncelli and Adelson (2002). A traditional intersection of constraints approach is enhanced by introducing an individual's decision uncertainty and priors into the process of motion computation. The 3D FORMOTION model can be viewed as the brain's way of using normalized patterns of form and motion activities as "real-time probabilities" that work together to contextually overcome uncertainty. Various properties of the 3D FORMOTION model receptive fields can be viewed as the outcome of developmental processes that are sensitive to the statistics of real-world scenes (cf., Grossberg and Swaminathan, 2004 and Grossberg and Williamson, 2001), and in this sense embody probabilistic constraints on model interactions. It should also be noted that any filtering operation, such as the model short-range and long-range filters, may be interpreted as a prior (namely, the current neural activity) multiplied by a conditional probability (namely, the filter connection strength to the target cell). Likewise, a contrast-enhancing competitive interaction that responds to such a filter may be viewed as a maximization operation. These insights have been known in the neural modeling literature for thirty years (e.g., Grossberg, 1978). However, as Figure 6 and 7 and the model equations in Appendix A show, such local processes do not, in themselves, embody the design constraints that lead to the emergent computational intelligence of an entire neural system.

Model parameters. The 3D FORMOTION model is more directly tied to primate neurophysiology than purely functional (e.g., Bayesian) models, but it provides a more lumped description of cell and network dynamics than, for example, multi-compartmental models of single neurons that include a large number of ionic conductances in each cell. Including all such factors would increase the number of parameters, and the run times, in our model many-fold with no gain in perceptual insight.

Because we used such a reduced parameter space, it is not possible to select model parameters based on data concerning individual cell firing rates as recorded in V1, V2 or other areas of visual cortex. The particular parameter values presently employed (given in Appendix A) can, however, be chosen in a robust parameter range without qualitatively changing the perceptual phenomena that the model can explain. Because the model is robust to changes in many parameters, it is compatible with previous motion models on which it builds, such as those of Baloch and Grossberg, (1997), Baloch et al. (1998), Chey et al., (1997, 1998) and Grossberg et al. (2001). While a full parameter comparison is given in Table 2 of Appendix A, four meaningful changes can here be noted: (1) Because the present simulations employ a higher complexity of motion signals in simulated displays and shorter spans of simulated time, the balance of excitation and opponent direction inhibition has changed: C_3, K_3, C_4, and K_4 at motion Level 2, equations (A7) and (A8); C_6 at motion Level 4, equation (A11); D_8, C_9, and D_9 at the motion Level 5 (MT and MST), equations (A16) and (A21). (2) The size of spatial kernels has been changed to reflect a different size of the display. A more fully developed model would include multiple scales of motion processing; here the optimal one was chosen for simplicity: σ_x and σ_y at the motion Level 5, equation (A16). (3) New mechanisms such as form boundary

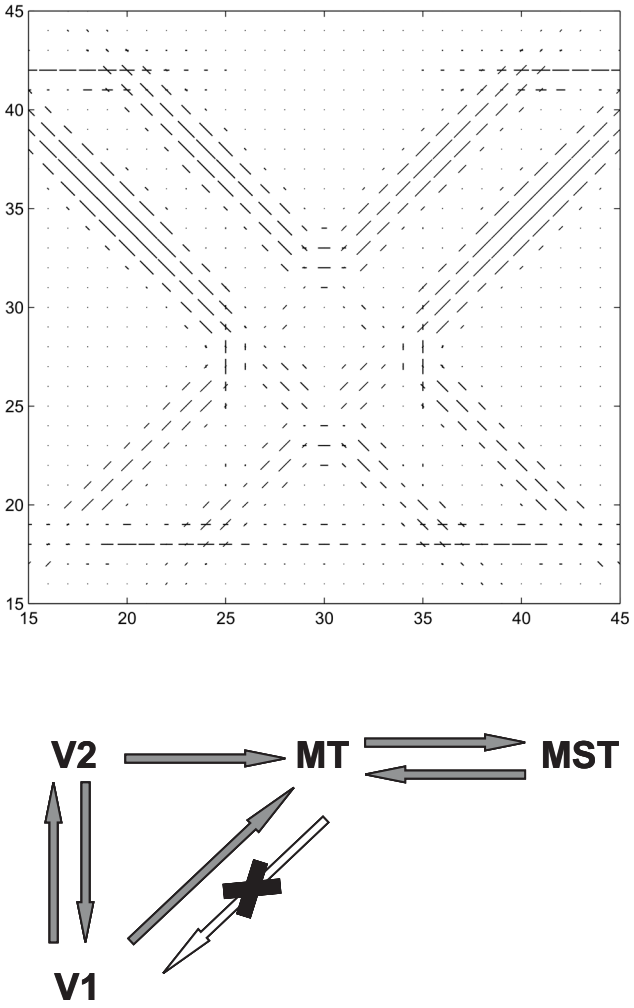

Figure 17. Chopsticks with invisible occluders and simulated model lesion. Boundary computation (bipole output, Form Level 5, Equation (A38)) without feedback from MT-V1. Orientation and length of short individual lines represent the orientation and strength of bipole cell activations at each position. Neither boundary can win, as the effects of attentional selection in the MT-MST loop cannot propagate to the form system via V1. Compare with Figure 13.

selection of motion signals via a V2-to-MT interaction required introduction of new layers in the model, such as motion Level 5, equation (A14). (4) Thresholds for motion signals at both the short-range (V1) and long-range (MT) motion filter levels prevent "leakage" of motion signals into the depth of static occluders: θ_1 and θ_2 at motion Level 3, equation (A9) and θ at motion Level 5 (MT), equation (A17).

A more intuitive way of understanding model interactions than screening individual parameters for a multi-layer system with feedback such as 3D FORMOTION is to examine the

model's performance when particular connections are "lesioned". For example, Figure 17 illustrates that the model is unable to separate chopsticks boundaries in depth in the invisible occluder case *without* feedback from MT to V1. This result is similar to initial simulation frames for a non-"lesioned" network (see Figure 13A). Without the breaking of symmetry afforded by a momentary attentional gain fluctuation that favors motion signals for one or the other chopstick, neither boundary representation can "win" and claim the near depth in the form system.

Figure 18 illustrates MT activity in the absence of MST feedback for the case of chopsticks with visible occluders. Unlike in Figure 14, motion integration and coherent grouping and selection is incomplete: multiple motion signals, many of them spurious, are found in the farther depth.

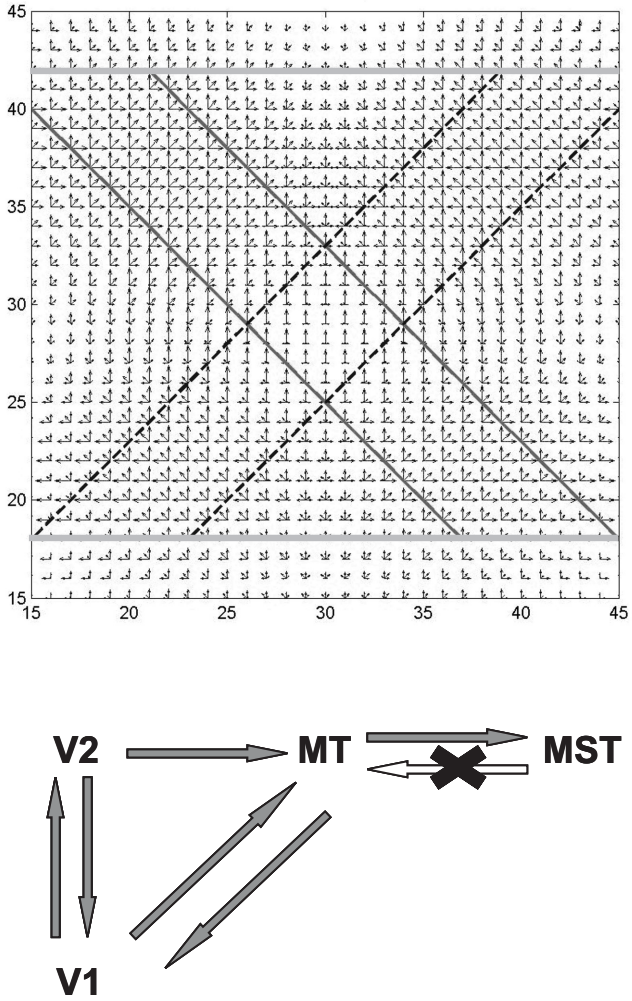

Figure 18. Chopsticks with visible occluders and simulated model lesion. MT activity (Motion Level 5, Equation (A16)) without MST feedback. Motion integration is incomplete, and multiple motion direction signals are found at the farther depth, D2. Compare with Figure 14B. Direction and length of individual arrows represent the direction and strength of MT cell activation at each position.

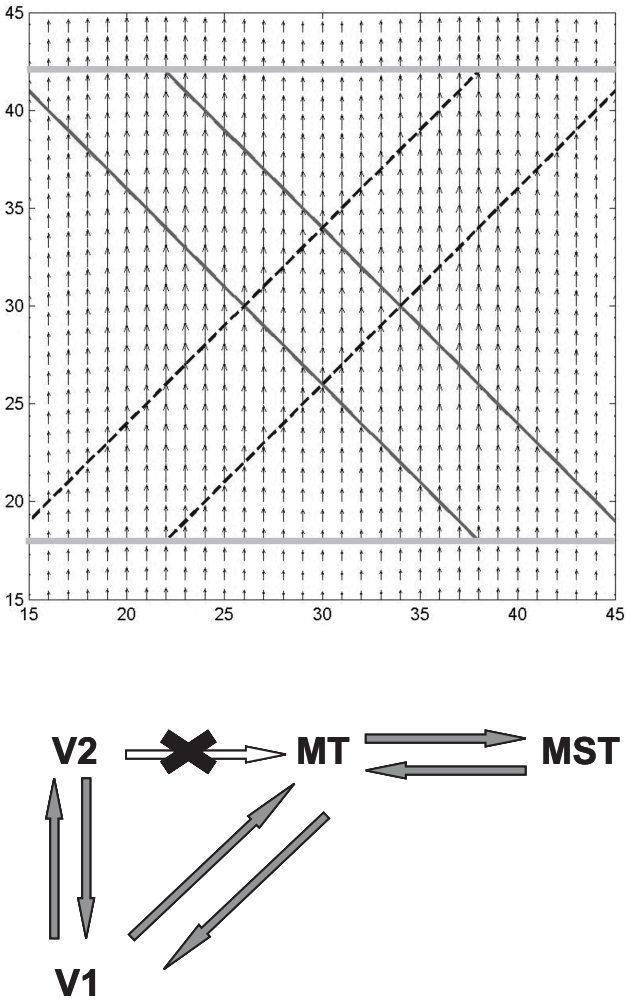

Figure 19. Chopsticks with visible occluders and simulated model lesion. MT activity (Motion Level 5, Equation (A16)) without V2-MT boundary selection. Nothing prevents unwanted motion signals in the closer depth, D1. Compare with Figure 14A. Direction and length of individual arrows represent the direction and strength of MT cell activation at each position.

Finally, Figure 19 illustrates MT activity in the absence of V2-to-MT boundary selection. In this case nothing prevents the occurrence of unwanted motion signals in the nearer depth. Compare this result with the perceptually correct lack of significant motion signals in the depth of the static occluder in the non-"lesioned" network in Figure 14A. While this paper does not explore all "lesion" possibilities in detail, results of the three types of model lesions described above can be experimentally tested *in vivo* by cooling, inhibitory agonist injection, or TMS stimulation. As noted above, the 3D FORMOTION prediction of the loss of depth selectivity in MT while preserving motion computations agrees well with the recent data on changes in activity of MT cells due to V2/V3 cooling (Ponce. Lomber and Born, 2006).

Conclusions: Mechanisms for Interaction of Form and Motion streams. One of the most important components of the 3D FORMOTION model is interaction between form and motion processing. Form and motion processing streams in the visual cortex are traditionally considered as separate from each other (Mishkin, Ungerleider and Macko, 1983). Separation starts at the retinal level. Lesion data seem to support the separation idea: Lesions of the parvocellular, or P-pathway, do not affect performance in pure motion tasks; lesions of the magnocellular, or M-pathway, do not affect color or fine spatial frequency sensitivity (Schiller and Logothetis, 1990). However, when more complicated motion scenes are considered, independence of the two pathways is questionable, as in the Lorenceau-Alais, chopsticks, and gelatinous ellipse displays.

3D FORMOTION uses interaction of form and motion streams to explain several perceptual phenomena. First, motion signals can change based on the occlusion information present in the display. For example, the difference between the motion of extrinsic and intrinsic terminators explains chopstick displays and some of the Lorenceau-Alais displays. Previously, it was suggested that FACADE figure-ground separation would provide a basis for such a distinction. However, separation of boundaries in depth does not happen until V2, or at least the upper layers of V1. Here we suggest that some difference between extrinsic and intrinsic terminators can already be detected in the input layers of V1, and it is established in part by adaptation to static boundaries. Electrophysiological recordings of V1 cell activity in response to a similar display, the diagonal grating with and without horizontal occluders (Pack et al., 2004), can be interpreted as a support to the adaptation hypothesis. These data are also consistent with properties of the model feedback from the V2 figure-ground separation mechanisms to the V1 motion stream. Because these authors did not study temporal dynamics of the suppression along occluders, or vary other parameters affecting depth order of the grating and horizontal occluders, based on their data it is hard to distinguish between feedforward and feedback mechanisms.

Second, 3D FORMOTION explains how the motions of two overlapping objects are separated in the MT-MST network. The projection from V1 to MT is unlikely to carry depth-selective signals (Movshon and Newsome, 1996). However, Palanca and DeAngelis (2003) have shown that cells in MT have disparity tuning even in the absence of motion. V2 cells appear to participate in figure-ground separation (Bakin, Nakayama and Gilbert, 2000, Zhou, Friedman and von der Heydt, 2000). The 3D FORMOTION model predicts that the V2 pale stripe projection to MT can carry occlusion information necessary to resolve the motion of different surfaces in depth. Such an on-center off-surround projection of depth-separated boundaries from V2 to the motion stream can also help to explain the absence of motion in the near depth of chopsticks (or any other) display with visible occluders. Occluder boundaries represented in the near depth plane would select relatively weak "extrinsic" motion signals along them and suppress motion signals anywhere else at that depth. This mechanism predicts that a proportion of cells in MT representing closer depths will be suppressed when occluder boundaries are presented. While neuronal recordings where either disparity-defined (Duncan et al., 2000) or contrast-defined (Pack et al., 2004) occluders were presented do not offer such evidence, protocols used in these studies did not include a control case of motion presented without occluders. Because only motion-sensitive cells are usually selected for recordings, the cell populations that are suppressed by form boundaries would be easy to overlook.

The 3D FORMOTION model makes specific predictions about the laminar distribution of Form-Motion interaction properties of MT cells. MT input cells modulated by localized V2 boundaries (Equation A14) are predicted to show a strong activation at the boundary positions and a weak one in empty spaces between boundaries. On the other hand, long horizontal

connections in superficial layers of MT are suggested to carry a motion grouping function (Equation A17) that renders cells less selective to a specific boundary position. This raises the question of why there seems to be an absence of a perceived motion signal in the intervening spaces between visible features, as happened in the case of induced motion (Duncker, 1929/1937). More selective motion binding to boundary positions may be due to replication of the V2-to-MT selection mechanism at a higher stage of processing than MT. Such a process is not implemented in this article. Another possibility is that not all active cells can carry a conscious percept of a perceptual quality. For example, if resonant activities at layer 4 become conscious, but not activities of cells in the long-range motion integration stage in layers 2/3, then no additional circuitry would be needed to explain the binding of motion percepts to emergent boundaries.

Other Form-Motion interaction phenomena can be explained by feedback projections between cortical areas. Different motion signals coexisting in the image can create a motion-defined boundary (separation in 2D plane) or two motion planes (separation in depth). This suggests that projections from the motion system go to the form boundary/surface processing system. Such a projection from MT to V1 was used in the present model to explain the perceived separation of chopsticks in depth in the invisible occluder case. Neurophysiological studies of the function of the MT-to-V1 projection (Movshon, and Newsome, 1996, Jones et al., 2001) used either microstimulation or microinjection techniques in the context of simple local motion displays. The effect of the feedback projection was often excitatory, sometimes inhibitory, but its overall function was not clear. We predict that it is realized by a modulatory on-center, off-surround network, much like in the MST-to-MT feedback pathway, and other attentional top-down circuits within the form processing stream (Grossberg, 1999; Raizada and Grossberg, 2003). Model predictions (Figure 17) can be tested using simultaneous recordings (from V1 or V2 cells) and inactivation of motion feedback areas (MT) using complex motion displays; for example, the chopsticks display that was used in our modeling study.

Projections from the motion to the form stream can also distort boundaries of objects under certain conditions, as in the case of gelatinous ellipses. In this article we show only the result of computations in the motion stream: a tangential motion in the case of the rigid ellipse, and radial motion in the case of non-rigid ellipse. Tangential and radial biases are consistent with rotation or deformation, respectively. A role for a motion-to-form projection in the distortion of boundary positions was explored in a follow-up of the current model (Berzhanskaya et al., 2004). Fu et al. (2004) demonstrated a motion-dependent shift of V1 receptive fields. Psychophysical experiments using TMS stimulation indicate the importance of MT-to-V1 connections for motion detection and a perceived position shift, albeit in a different paradigm (Silvanto, Lavie and Walsh, 2005; McGraw, Walsh and Barret, 2004). Further neurophysiological experiments are needed to test if MT-to-V1 projections are responsible for this shift and for a deformation percept in the case of non-rigid ellipse

One important difference between the form and motion systems is their difference in timing. In particular, the timing of boundary completion is sometimes slow because it may involve feedback and competition between different depth planes. There are also latency differences between parvocellular and magnocellular streams. The motion signal to MT is very quick with a latency of 40 ms, compared to more than 50 ms in orientation-selective simple and complex cells in V1 (Bullier, 2001; Bair, Cavanaugh, Smith and Movshon, 2002). While adaptation mechanisms resulting in the intrinsic/extrinsic terminator distinction are feedforward and quick, boundary selection mechanisms require an additional stage of cortical processing and

are slower. On the other hand, motion signals, even in a simple moving line display, suffer from the aperture problem. In the visible occluder case of the chopsticks display, the 3D FORMOTION model predicts that, initially, the direction of motion in both depth representations of MT corresponds to an ambiguous motion signal, and that the correct motion signal develops through time. With time, boundary suppression through the V2-to-MT projection starts to inhibit the motion signal in the near depth plane, concurrently with the development of the correct motion signal in the farther depth plane. This effect could be noticeable in the depth-modulated barberpole illusion, as in Duncan et al. (2000), if experiments are modified to afford an analysis of timing of motion-sensitive cells relative to boundary onset. Pack, Berezovskii, and Born (2001) did demonstrate a switch from ambiguous to veridical direction of motion over a period of 50-70ms as detected in the response of certain MT cells to a modified barberpole illusion. The effect of suppression of motion in the corresponding depth remains to be shown. On the other hand, a longer time-scale phenomena, such as alternation between coherent and transparent plaid motion with a characteristic time of 1-5 min (Hupe and Rubin, 2003), can be explained by adaptation to an active motion direction. Such an adaptation mechanism has been used at the MT-MST stage to explain related data about plaid adaptation. See Chey, Grossberg,and Mingolla (1997, Section 7C) and Grossberg, Mingolla, and Viswanathan (2001, Section 3.10). An adaptation to a coherent plaid direction of motion would allow other strong directions (component motion) to win and facilitate separation of motion in depth, as in the case of two chopsticks with invisible occluders simulated in this paper.

The 3D FORMOTION model explanations are consistent with those of many other motion data by earlier versions of the model (Baloch and Grossberg, 1998; Chey et al., 1997; Francis and Grossberg, 1996; Grossberg et al., 2001). The same mechanisms can be also applied to illusory boundaries from motion (Anderson and Barth, 2000), aperture discontinuity (Palmer and Kellman, 2001), flash lag and flash-drag effects (Nijhawan, 1994; Whitney and Cavanagh, 2000), and motion induction/motion capture effects (Murakami 1999). Some of these issues are addressed in a follow-up of the current model (Berzhanskaya et al., 2004).

Acknowledgements

J. Berzhanskaya was supported in part by the Air Force Office of Scientific Research (AFOSR F49620-01-1-0397), the National Geospatial-Intelligence Agency (NMA201-01-1-2016), the National Science Foundation (NSF BCS-0235398), and the Office of Naval Research (ONR N00014-95-1-0409 and ONR N00014-01-1-0624).

S. Grossberg was supported in part by the National Science Foundation (NSF SBE-0354378) and the Office of Naval Research (ONR N00014-01-1-0624).

E. Mingolla was supported in part by the National Geospatial-Intelligence Agency (NMA201-01-1-2016), the National Science Foundation (NSF BCS-02-35398 and NSF SBE-0354378), and the Office of Naval Research (ONR N00014-01-1-0624).

Appendix A: 3D FORMOTION equations, parameters and implementation

All stages of the model, except simple cells in the form system (Equation A23) were numerically integrated using a 4th order Runge-Kutta method with a fixed integration step. The activity of simple cells was computed at equilibrium. Each layer, including the input, was represented by a 60x60 matrix for each combination of attributes used at the given layer. For example, for the motion system, if there were 2 spatial scales, there were (2x8) cells sensitive to different combinations of scale and direction at each point of the matrix. For the scale-sensitive form system cells, there were (2x4) scale and orientation cells at each point in the image. For visual clarity, figures depict activities of the central part of the corresponding layers (about 30x30 cells), where most of the input motion was generated.

I. Motion system

All motion sequences are given to the network as series of static 2D frames representing black-and-white image snapshots at the consecutive moments of time. In both form and motion systems, inputs are not separated in depth; i.e., both occluder and occluded objects exist in the same image plane. Activities at each layer (y_n) are results of computation in a dynamical system, where the rate of activity change is proportional to some function *f* of this layer's activities, inputs *I* and, sometimes, feedback *F*. Dynamics can be described in a general form as:

$$\frac{dy_n}{dt} = A_n f(y_n, I, F), \qquad \text{(A1)}$$

where A_n scales how fast y_n changes. High values of A_n result in fast dynamics, while low values of A_n result in slow dynamics. Outputs of all stages are rectified: $Y_n = [y_n]^+ = \max(y_n, 0)$. All model equations are membrane equations:

$$C_m \frac{dV}{dt} = -[V - E_{excit}]g_{excit} - [V - E_{inhib}]g_{inhib} - [V - E_{leak}]g_{leak} \qquad \text{(A2)}$$

In this equation, g_{excit} and g_{inhib} represent the total inputs from excitatory and inhibitory neurons synapsing on the cell; g_{leak} is a leakage conductance. Parameters E_{excit} , $E_{inhib,}$ and E_{leak} are reversal potentials for excitatory, inhibitory and leakage conductances, respectively. All conductances contribute to the divisive normalization of the membrane potential, V, as shown by equilibrium solution for V:

$$V = \frac{(E_{excit}g_{excit} + E_{inhib}g_{inhib} + E_{leak}g_{leak})}{(g_{excit} + g_{inhib} + g_{leak})} \qquad \text{(A3)}$$

(Grossberg, 1973, 1980; Grossberg and Raizada, 2000). Reversal potentials in the following simulations were (for simplicity) set to $E_{excit} = 1$, $E_{inhib} = -1$, and $E_{leak} = 0$ (unless noted otherwise). When the reversal potential of the inhibitory channel, E_{inhib}, is close to the resting potential, the inhibitory effect is pure "shunting"; i.e., decreasing effect of excitation only through an increased membrane conductance. It balances excitatory inputs and prevents network activities from saturating. In the equations where saturation effects are not possible (for example A9), the shunting term was not used.

Depending on a layer's functionality, activities at each position (*i,j*) are represented as x_{ij}^p, where $p \in \{1,2\}$ indicates whether the cell (population) belongs to an ON or OFF stream; or as

x_{ij}^{d}, where $d \in \{1,...,8\}$ indicates directional preference within a single spatial scale; or else as x_{ij}^{ds} where $d \in \{1,...,8\}$ indicates motion directional preference, and $s \in \{1,2\}$ indicates spatial scale.

Level 1: Input. Motion processing starts from the input layer of V1 ($4C_\alpha$). Previous models (Baloch et al., 1997) analyzed how LGN ON and OFF cell streams interact to create boundaries from a 2D image. They demonstrated that in static images ON cells within on-center off-surround networks, and OFF cells within off-center on-surround networks create thin boundaries on the edges of the object. Boundaries at the leading edge of a moving bright bar are represented mainly by the ON stream, while boundaries at the trailing edge are represented mainly by the OFF stream. Based on the results of Baloch et al. (1997), a simplified input I_{ij}^{p} to the visual cortex was represented by 2-cell wide boundaries in two separated ON and OFF channels. This simplification was motivated by the fact that we used simple black-and-white images. The boundary on the leading edge of the object was represented by the ON channel, and the boundary on the trailing edge by the OFF channel. No interactions between ON and OFF channels were simulated.

Level 2: Transient cells. At the first stage of V1, non-directional transient cell activities b_{ij} are computed as a sum of ON ($p = 1$) and OFF ($p = 2$) channels:

$$b_{ij} = \sum_{p} x_{ij}^{p} z_{ij}^{p}, \tag{A4}$$

where input cell activities, x_{ij}^{p}, perform leaky integration on their inputs I_{ij}^{p}:

$$\frac{dx_{ij}^{p}}{dt} = A_1 \left(-B_1 x_{ij}^{p} + \left(C_1 - x_{ij}^{p} \right) I_{ij}^{p} \right). \tag{A5}$$

Non-zero activation x_{ij}^{p} results in slow adaptation of a habituative transmitter gate z_{ij}^{p}:

$$\frac{dz_{ij}^{p}}{dt} = A_2 \left(1 - z_{ij}^{p} - K_2 x_{ij}^{p} z_{ij}^{p} \right) \tag{A6}$$

(Abbott et al., 1997; Grossberg, 1980). In (A5), $A_1 B_1 x_{ij}^{p}$ is the rate of passive decay and C_1 is a maximum activity x_{ij}^{p} can reach. For non-zero inputs I_{ij}^{p}, x_{ij}^{p} approaches C_1 with a rate proportional to ($C_1 - x_{ij}^{p}$) and decays with the rate proportional to $B_1 x_{ij}^{p}$. When a nonzero input x_{ij}^{p} is presented, z_{ij}^{p} adapts with the rate of $A_2 K_2 x_{ij}^{p}$ in (A6). When the input returns to 0, z_{ij}^{p} recovers to 1 at the rate A_2. The parameters used in Level 2 simulations are: $A_1 = 10$, $B_1 = 3$, $C_1 = 1$, $A_2 = 0.01$, and $K_2 = 20$.

Input activity x_{ij}^{p} combined with transmitter gate z_{ij}^{p} results in transient non-directional cell activities b_{ij} that model activity of the non-directionally selective cells in layers $4C_a$ with circular receptive fields (Livingstone and Hubel, 1984). ON and OFF inputs summate at this stage. For visual inputs with a short dwell time, such as moving boundaries, activities b_{ij} respond well. A static input, on the other hand, produces only a weak response after an initial presentation period, because of the habituation (Muller, Metha, Krauskopf, and Lennie, 2001).

The next two cell layers provide a directional selectivity mechanism that can retain its sensitivity in response to variable speed inputs (Chey et al., 1997). As noted above, index d

denotes the directional preference of a given cell. First, directional interneuron activities c_{ij}^{d} integrate transient cell inputs b_{ij}:

$$\frac{dc_{ij}^{d}}{dt} = A_3\left(-B_3 c_{ij}^{d} + C_3 b_{ij} - K_3\left[c_{XY}^{D}\right]^{+}\right). \quad \text{(A7)}$$

A directional inhibitory interneuron c_{ij}^{d} receives excitatory input from a transient non-directional cell activity b_{ij} at the same position, and suppression from directional interneuron c_{XY}^{D} of opposite direction preference D at the position (X,Y) offset by 1 cell in the direction d. For example, for the direction of motion 45°, $X = i+1$, $Y = j+1$, and $D = 135°$.

Activity c_{ij}^{d} increases proportionally to input b_{ij} with coefficient A_3C_3 and passively decays to zero with rate $A_3B_3 c_{ij}^{d}$. The strength of opponent inhibition is $K_3\left[\boldsymbol{c}_{\boldsymbol{XY}}^{\boldsymbol{D}}\right]^{+}$. Inhibition is stronger than excitation and "vetoes" a directional signal if the stimulus arrives from the null direction. Thus, a bar arriving from the left into the rightward directional interneuron receptive field would activate it well; while a bar arriving from the right would inhibit it even if activation b_{ij} is non-zero. The parameters are: $A_3 = 5$, $B_3 = 2$, $C_3 = 0.5$, and $K_3 = 20$.

Directional transient cell activities e_{ij}^{d} at the next level combine transient input b_{ij} with inhibitory interneuron activity c_{ij}^{d}. Their dynamics are similar to those of c_{ij}^{d}:

$$\frac{de_{ij}^{d}}{dt} = A_4\left(-B_4 e_{ij}^{d} + C_4 b_{ij} - K_4\left[c_{XY}^{D}\right]^{+}\right). \quad \text{(A8)}$$

Activity e_{ij}^{d} increases proportionally to transient input b_{ij}, passively decays with the fixed rate, and is inhibited by an inhibitory interneuron tuned to the opponent direction. The parameters are: $A_4 = 30$, $B_4 = 1$, $C_4 = 0.5$, and $K_3 = 20$. Computation at Level 2 results in multiple directions activated in response to a moving line, which is consistent with the ambiguity caused by the aperture problem due to the limited size of V1 receptive fields.

Level 3: Short-range motion filter. Short-range anisotropic filter activities, f_{ij}^{ds}, accumulate motion in each direction d:

$$\frac{df_{ij}^{ds}}{dt} = A_5\left(-f_{ij}^{ds} + \sum_{XY} E_{XY}^{d} G_{ijXY}^{ds}\right). \quad \text{(A9)}$$

Here E_{ij}^{d} is the rectified output of e_{ij}^{d} from Level 2, and G_{ijXY}^{ds} is a Gaussian receptive field that depends on both direction d and scale s:

$$G_{ijXY}^{ds} = G\exp\left(-0.5\left(\left(\frac{X-i}{\sigma_x^{s}}\right)^2 + \left(\frac{Y-i}{\sigma_y^{s}}\right)^2\right)\right). \quad \text{(A10)}$$

Scale s determines a receptive field size, and therefore the extent of spatiotemporal integration of lower-level motion signals. Larger receptive fields respond selectively to larger speeds, smaller receptive fields to smaller speeds; cf., Chey et al. (1998). While in our simulations speed did not vary much, in more motion-rich environments speed-depth correlations can help to assign an approximate depth order to the moving objects. The kernel G_{ijXY}^{ds} is elongated in the direction of

motion. For a horizontal motion direction, the kernel has σ_x^s = 1.5, σ_y^s = 0.5 for s=1; σ_x^s = 2.5, σ_y^s = 0.5 for s = 2; G = 0.15. Kernels for other directions are derived by a rotation which aligns the major kernel axis with the preferred direction of motion. Output of the short-range filter is thresholded and rectified, $F_{ij}^{ds} = \left[f_{ij}^{ds} - \theta_s\right]^+$, with threshold θ_1 = 0.04, θ_2 = 0.08. Self-similar scale-specific thresholds provide different speed sensitivity for two spatial scales. If thresholds for two scales were the same, the larger scale would be always activated more strongly. With the larger threshold it prefers larger speeds. The full simulation of speed sensitivity was performed in a similar system by Chey et al. (1997). The value of constant A_5 = 50.

Level 4: Spatial competition and opponent direction inhibition. The next cell layer activities, h_{ij}^{ds}, combine spatial competition within one motion direction across the area determined by the kernel K_{ijXY}^{ds} with inhibition from opponent direction cells F_{ij}^{Ds} in the same spatial position. A membrane, or shunting, equation combines these effects:

$$\frac{dh_{ij}^{ds}}{dt} = A_6\left(-h_{ij}^{ds} + \left(1 - h_{ij}^{ds}\right)\sum_{XY} F_{XY}^{ds} J_{ijXY}^{ds} - \left(0.1 + h_{ij}^{ds}\right)\left[C_6 \sum_{XY} F_{XY}^{ds} K_{ijXY}^{ds} + D_6 F_{ij}^{Ds}\right]\right). \quad \text{(A11)}$$

Rectified activities, F_{ij}^{ds}, from Level 3 define the spatial competition through the excitatory Gaussian kernel J_{ijXY}^{ds}, which is spatially anisotropic with σ_x = 2.5 and σ_y = 0.5 (for horizontal motion):

$$J_{ijXY}^{ds} = \frac{J}{2\pi\sigma_x\sigma_y}\exp\left(-0.5\left(\frac{(X-i)^2}{\sigma_x^{\,2}} + \frac{(Y-j)^2}{\sigma_y^{\,2}}\right)\right), \quad \text{(A12)}$$

and the inhibitory kernel K_{ijXY}^{ds}, which is isotropic with σ = 4:

$$K_{ijXY}^{ds} = \frac{K}{2\pi\sigma^2}\exp\left(-0.5\left(\frac{(X-i)^2 + (Y-j)^2}{\sigma^2}\right)\right). \quad \text{(A13)}$$

The center of inhibitory kernel K_{ijXY}^{ds} is offset from the (i,j) position by one cell in the direction opposite to the cell preferred direction d. This arrangement results in inhibition trailing excitation. The strength of spatial competition is determined by parameter C_6, and that of opponent inhibition by D_6. Parameters are: A_6 = 50, C_6 = 5, D_6 = 100, J = 2, and K = 2. $\boldsymbol{D}$ is opposite to d.

It usually takes few frames of motion to accumulate and accurately compute motion signals through the Level 2-4 mechanisms (Equations A4-A13). However, a motion span (maximal displacement in one direction) of the Lorenceau-Alais displays is small. The radius of rotation and the motion span there are limited by the geometry of the input; in particular, corners of the shape that provide unambiguous motion signals are not visible. To accumulate enough information for the motion mechanisms to adequately sample the moving stimulus, one may increase the size of the network by supersampling and scale the motion sequence correspondingly. For example, a 3-pixel sequence of motion in one direction becomes a 9-pixel sequence (scaling by a factor of three). In order to keep the simulation times reasonable, this scaling was done only up to Level 4 (see Figure 7, layers 4C-4B, and Equations A4-A13). Furthermore, due to memory restrictions, displays were computed piece-wise: four segments of each shape were processed by a 60 x 60 network each. Output activities at Level 4 were then

subsampled by a factor of 3, in order to compensate for the previous supersampling, and combined into one 60 x 60 display at Level 5 (Equation A14). A supersampled 9-pixel motion sequence thus becomes a subsampled 3-pixel sequence, thereby returning to the original cellular dimensions, but motion signals are more thoroughly processed due to the finer scale at the input levels. Piece-wise simplification was possible because four segments of an individual Lorenceau-Alais shape are separated in space and do not interact with each other at the spatial scale of Level 2-4 computations. When interactions between segments become essential (Level 5 and later), activities are combined. Computations for Lorenceau-Alais used the same parameters as for other displays.

Level 5: Formotion capture and long-range filter. Rectified motion output signals, H_{ij}^{ds}, from V1 (model Level 4) are selected by form boundary signals, $\tilde{z}_{ij}^{s}$, from V2 in the input layers 4 and 6 of MT. The activities, q_{ij}^{ds}, of these MT cells combine motion and boundary signals via a membrane equation:

$$\frac{dq_{ij}^{ds}}{dt} = A_7\left(-q_{ij}^{ds} + \left(1-q_{ij}^{ds}\right)H_{ij}^{ds}\left(K_e + K_z\tilde{z}_{ij}^{s}\right) - K_b\left(1+q_{ij}^{ds}\right)\sum_{XY}\tilde{z}_{XY}^{s}I_{ijXY}^{s}\right). \quad \text{(A14)}$$

In (A14), an input from the V1 motion stream $K_e H_{ij}^{ds}$ is positively modulated by boundaries $K_z\tilde{z}_{ij}^{s}$ in the excitatory term of the equation (A14). In addition, boundaries inhibit unmatched motion signals via term $\sum_{XY}\tilde{z}_{iXY}^{s}I_{ijXY}^{s}$. This modulatory on-center off-surround network allows boundaries to select motion signals at their positions and corresponding depths. Parameter K_e determines the strength of feedforward inputs H_{ij}^{ds}, and K_z the strength of modulation by V2 boundaries. The V2 boundary projection to MT is stronger than the bottom-up motion projection; that is, $K_e << K_z$. The strength of the inhibitory effect of V2 boundaries $\tilde{z}_{ij}^{s}$ is proportional to the coefficient K_b, and its spatial reach is determined by inhibitory Gaussian kernel $\boldsymbol{I_{ijXY}^{s}}$:

$$I_{ijXY}^{ds} = \frac{I}{2\pi\sigma^2}\exp\left(-0.5\left(\frac{(X-i)^2+(Y-j)^2}{\sigma^2}\right)\right). \quad \text{(A15)}$$

When no boundary is provided and $\tilde{z}_{ij}^{s}$ is 0 everywhere (for example, the parvocellular stream is inactivated), motion signals can still activate MT via the term $K_e H_{ij}^{ds}$ in (A14). In this case, no inhibition is present as well. In the presence of boundary input, motion signals at the boundary positions are strong, whereas those outside of the boundary position are suppressed. Activity $\tilde{z}_{ij}^{s}$ in (A14) codes an idealized 1-cell wide boundary simulating output of V2. It simplifies boundaries z_{ij}^{s} separated in depth by the form system (Equation A38), positioned at the locations of input boundaries I_{ij} (Equation A4). Parameter $s = 1$ corresponds to the near depth, $s = 2$, to the far depth. The parameters are: $A_7 = 100$, $K_e = 1$, $K_z = 10$, $K_b = 0.12$, $I = 0.1$, and $\sigma = 6$.

Next, MT cell activities, m_{ij}^{ds}, in layer 2/3 receive MT signals, N_{ij}^{ds}, from layer 4 via a long-range filter and top-down matching signals, T_{ij}^{ds}, from MST:

$$\frac{dm_{ij}^{ds}}{dt} = A_8\left(-m_{ij}^{ds} + \left(1 - m_{ij}^{ds}\right) N_{ij}^{ds} - D_8\left(1 + m_{ij}^{ds}\right) \sum_{e,XY} w^{de} \left[T_{XY}^{es} \right]^+ P_{ijXY}^{s} \right). \quad \text{(A16)}$$

To compute the long-range filter inputs, N_{ij}^{ds}, the MT input activities, q_{ij}^{ds}, are rectified ($Q_{ij}^{ds} = \left[q_{ij}^{ds}\right]^+$), squared, and passed through an anisotropic filter L_{ijXY}^{ds}, thresholded, and rectified again:

$$N_{ij}^{ds} = \left[\sum_{XY} \left(Q_{XY}^{ds}\right)^2 L_{ijXY}^{ds} - \theta_n \right]^+ . \quad \text{(A17)}$$

In (A17), L_{ijXY}^{ds} is an anisotropic Gaussian kernel:

$$L_{ijXY}^{ds} = \frac{L}{2\pi\sigma_x\sigma_y} \exp\left(-0.5\left(\left(\frac{X-i}{\sigma_x} \right)^2 + \left(\frac{Y-i}{\sigma_y} \right)^2 \right) \right) \quad \text{(A18)}$$

that is elongated in the direction of preferred motion. For the horizontal motion direction, for example, σ_x = 10, σ_y = 4. L = 20, and θ = 0.03. Kernels for other directions are derived by rotation, for example to compute 45° kernel, the horizontal kernel is rotated 45° counterclockwise.

This long-range anisotropic Gaussian filter accumulates motion in its preferred direction over time and space. The anatomical basis for such integration can be provided by long-range horizontal projections in layers 2/3 of MT. The squaring operation gives higher preference to larger signals, which leads to winner-take-all dynamics in competitive recurrent networks (Grossberg, 1973, 1988).

The strength of MST feedback is determined by coefficient D_8. Its spatial extent is determined by the isotropic kernel P_{ijXY}^{ds}:

$$P_{ijXY}^{s} = \frac{1}{2\pi\sigma^2} \exp\left(-0.5\left(\frac{(X-i)^2 + (Y-j)^2}{\sigma^2} \right) \right) \quad \text{(A19)}$$

with σ = 8. The suppression is from all directions except d. The inhibitory weight w^{de} between given direction d and another direction e is given by:

$$w^{de} = \begin{cases} 0, & e = d \\ 1, & e \neq d, e \neq D \\ 2, & e = D \end{cases} , \quad \text{(A20)}$$

where D is the direction opposite to d. Because excitatory input N_{ij}^{ds} is from the preferred direction, this asymmetric suppression effectively amplifies d and suppresses other motion directions. Motion from unambiguous feature-tracking signals propagates to ambiguous motion positions through the large kernel $\boldsymbol{P_{ijXY}^{s}}$. As in the case of the V2-to-MT and MT-to-V1 projections, MST-to-MT feedback is defined by a modulatory on-center, off-surround network. The parameters are: A_8 = 200 and D_8 = 5.

Level 6: Directional grouping and suppression in depth. The MT-MST circuit acts in a winner-take-all mode, selecting a single direction of motion at each point. MST activity $\boldsymbol{T_{ij}^{es}}$ is described by

$$\frac{dT_{ij}^{ds}}{dt} = A_9\left(-T_{ij}^{ds} + \left(1 - T_{ij}^{ds}\right) M_{ij}^{ds}\left(1 + O_{ij}^{ds}\right) - D_9 \sum_{D,XY} w^{dD}\left[T_{XY}^{es}\right]^{+} P_{ijXY}^{s} - C_9 \sum_{s<S} T_{ij}^{ds}\right), \qquad \text{(A21)}$$

where M_{ij}^{ds} is the rectified MT output $M_{ij}^{ds} = \left[m_{ij}^{ds}\right]^{+}$. Inhibition is provided by recurrent connections within MST. Its strength is determined by coefficient D_9, and its spatial extent by the kernel P_{ijXY}^{ds}. The weighting coefficient w^{de} and surround suppression kernel P_{ijXY}^{s} are the same as in equation (A16). MST also includes direction-specific suppression from the near depth (D1, s = 1) to the far depth (D2, s = 2), which is important for motion transparency simulations. If the motion in the direction d wins in D1, the same direction will be suppressed in D2. This allows the model to avoid a single motion direction being represented in both depths. In the case of transparent motion, suppression of one direction in a given depth would allow the other direction to win. The parameters are: A_9 = 400, C_9 = 1, and D_9 = 10.

MST can be modulated by attention via term O_{ij}^{ds} in Equation (A21). If attention is attracted by features in the near depth plane, this would help one motion direction to win in the near depth. Attention was used only in chopsticks simulations with invisible occluders to break the symmetry between competing motion signals from two chopsticks moving in opposite directions. Attention was applied as a single Gaussian "spot" in the near depth (s = 1) and rightward direction (d = 5):

$$O_{ij}^{ds} = A\exp\left(-0.5\left(\frac{(x_0 - i)^2 + (y_0 - j)^2}{\sigma^2}\right)\right). \qquad \text{(A22)}$$

In (A22), σ = 2, A = 0.05, and (x_0, y_0) are the spatial coordinates of the top-left moving chopstick end. This bias is similar to the one used in the case of transparent motion in Grossberg et al. (2001) and allows a single motion signal to win in D1.

II. Form system

A reduced version of the FACADE model was implemented as the form system in order to keep simulations manageable. Only part of the Boundary Contour System (BCS) was simulated. Binocular inputs were not considered. Filling-in was not simulated. After completion of object boundaries, it was assumed that filling-in processes and boundary competition in a full FACADE implementation would complete separation of occluded and occluding objects in different depths, as has been demonstrated in other articles; e.g., Cao and Grossberg (2005), Fang and Grossberg (2004), Grossberg and Swaminathan (2004), Grossberg and Yazdanbakhsh (2005), and Kelly and Grossberg (2000).

Depending on a layer's functionality, activities at each position (i,j) are represented as x_{ij}^{p}, where p indicates either whether a cell population belongs to ON or OFF streams, or whether it is an odd/even filter; or as x_{ij}^{rs}, where $r \in \{1,2,3,4\}$ indicates orientational preference, and $s \in \{1,2\}$ indicates the spatial scale or depth plane.

Level 1: Input. Input to V1, X_{ij}^{p}, corresponds to the input processing by LGN through circular center-surround receptive fields. As in the motion system, a simplified input $X_{ij}^{ON/OFF}$ was represented by 1-cell wide boundaries in two distinct ON and OFF channels. This simplification was motivated by our use of simple black-and-white images. No interaction between ON and OFF channels was simulated.

Level 2: Simple cells. Model simple cells respond to oriented contrasts in the image in a polarity-sensitive manner. Simple cell activities for orientation *d* and spatial scale s, S_{ij}^{rs}, are computed by convolution of V1 input $X_{ij}^{ON/OFF}$ with even-symmetric and odd-symmetric oriented filters $s_{ijpq}^{rs,even}$ and $s_{ijpq}^{rs,odd}$, respectively:

$$S_{ij}^{rs,ON/OFF,even} = \left\lfloor \sum_{pq} s_{ijpq}^{rs,even} X_{pq}^{ON/OFF} - \sum_{pq} s_{ijpq}^{rs,even} X_{pq}^{OFF/ON} \right\rfloor, \qquad \text{(A23)}$$

$$S_{ij}^{rs,ON/OFF,odd} = \left\lfloor \sum_{pq} s_{ijpq}^{rs,odd} X_{pq}^{ON/OFF} - \sum_{pq} s_{ijpq}^{rs,odd} X_{pq}^{OFF/ON} \right\rfloor, \qquad \text{(A24)}$$

where odd and even filters are given by

$$s_{ijpq}^{rs,odd} = \exp\left[-0.5\left(\frac{(i-p)^2}{\sigma_x^{s\,2}} + \frac{(j-q)^2}{\sigma_y^{s\,2}}\right)\right] - \exp\left[-0.5\left(\frac{(i-p-h)^2}{\sigma_x^{s\,2}} + \frac{(j-q)^2}{\sigma_y^{s\,2}}\right)\right] \qquad \text{(A25)}$$

and

$$s_{ijpq}^{rs,even} = k_1 \exp\left[-0.5\left(\frac{(i-p)^2}{\sigma_x^{s\,2}} + \frac{(j-q)^2}{\sigma_y^{s\,2}}\right)\right] - k_2 \exp\left[-0.5\left(\frac{(i-p-h)^2}{\sigma_x^{s\,2}} + \frac{(j-q)^2}{\sigma_y^{s\,2}}\right)\right]$$
$$- k_2 \exp\left[-0.5\left(\frac{(i-p+h)^2}{\sigma_x^{s\,2}} + \frac{(j-q)^2}{\sigma_y^{s\,2}}\right)\right]. \qquad \text{(A26)}$$

Four orientations (*r*) and two scales (*s*) were used. For the vertically oriented filter the parameters are: $\sigma_x^{r1} = \sigma_x^{r2} = 0.75$, $\sigma_y^{r1} = 1$, $\sigma_y^{r2} = 2.5$; $h = 1$, $k_1 = 1.6$, and $k_2 = 0.8$.

Level 3: Complex cells. Complex cells pool inputs from simple cells of the same orientation and opposite polarity. Complex cell activities c_{ij}^{rs} combine odd and even, and ON and OFF, inputs from simple cells:

$$G_{ij}^{rs} = \left|S_{ij}^{rs,even,ON} - S_{ij}^{rs,even,OFF}\right| + \left|S_{ij}^{rs,odd,ON} - S_{ij}^{rs,odd,OFF}\right|. \qquad \text{(A27)}$$

The activity of the complex cells is computed as:

$$\frac{dc_{ij}^{rs}}{dt} = A_2\left(-c_{ij}^{rs} + B_2\left(1-c_{ij}^{rs}\right)G_{ij}^{rs}\left(1+k_{ex}M_{ij}^{s}+k_{v2}\tilde{z}_{ij}^{s}\right) - D_2\left(1+c_{ij}^{rs}\right)\sum_{S\neq s} c_{ij}^{rs}\right). \qquad \text{(A28)}$$

The last term introduces competition between boundaries of the same orientation *r*, at each point (*i,j*), across scales *S* and *s*, thus allowing a given orientation boundary to be represented in only one depth.

Both V2 and MT modulate the response of V1 complex cells to the simple cell inputs, G_{ij}^{rs}, via the terms that multiply, and thus modulate, G_{ij}^{rs} in Equation (A28). Thus, the model predicts that MT feedback:

$$M_{ij}^{s} = \sum_{d} \left\lfloor M_{ij}^{ds} - \vartheta_{MT} \right\rfloor, \qquad \text{(A29)}$$

to layer 2/3 via apical dendrites in layer 1 of V1 (Callaway, 1998; Shipp and Zeki, 1989), where M_{ij}^{ds} is the rectified MT output $M_{ij}^{ds} = \left[m_{ij}^{ds}\right]^+$ computed in Equation (A16), can affect not

only the motion system but also complex cells in the form system (second term in Equation (A28)). This feedback is scale and depth specific, but not orientation or direction specific, and provides excitatory modulation only. MT-to-V1 feedback may also modulate layer 4 or V1, but this projection would have no effect on simulation of the targeted data.

The second feedback component of the second term in (A28) is provided by a 1-cell wide approximation $\tilde{z}_{ij}^{rs}$ of V2 boundaries z_{ij}^{rs} (see Equation (A28), and also (A38) and (A43) below). $\tilde{z}_{ij}^{rs}$ positions correspond to the boundary positions at the input level I_{ij}, and their value is 1 on a background of 0's. These boundaries provide a depth-specific bias to V1 complex cells, so if FACADE mechanisms assign a certain boundary to depth 1 at the level of V2, this boundary will be strengthened in scale 1 at the complex cell layer in V1, and will be weakened in scale 2. This mechanism helps to ensure that a given boundary is represented in one depth only. Parameters are: $A_2 = 2$, $B_2 = 1$, $k_{ex} = 25$, $\theta_{MT} = 0.15$, and $D_2 = 10$.

Level 4: Hypercomplex cells. The hypercomplex cell level has both spatial, y_{ij}^{rs}, and orientational, n_{ij}^{rs}, competition stages. This level models the process of end-stopping. It combines feedforward inputs from complex cells, C_{pq}^{rs}, through the on-center off-surround terms C_4 and E_4, respectively, with feedback inputs from bipole cells through the on-center off-surround terms C_7 and E_7, respectively. The activity, y^{rs}, at the spatial competition stage is described by:

$$\frac{dy_{ij}^{rs}}{dt} = A_3\left(-0.1y_{ij}^{rs} + (1-y_{ij}^{rs})\left[C_{4ij}^{rs} + C_{7ij}^{rs}\right] - (y_{ij}^{rs}+1)\left[E_{4ij}^{rs} + E_{7ij}^{rs}\right]\right). \qquad \text{(A30)}$$

The on-center, C_{4ij}^{rs}, and the off-surround, E_{4ij}^{rs}, inputs from complex cells obey:

$$C_{4ij}^{rs} = \frac{C}{2\pi\sigma_c^2}\sum_{pq}\exp\left[-0.5\left(\frac{(i-p)^2+(j-q)^2}{\sigma_c^2}\right)\right]C_{pq}^{rs}, \qquad \text{(A31)}$$

and

$$E_{4ij}^{rs} = \frac{E}{2\pi\sigma_s^2}\sum_{pq}\exp\left[-0.5\left(\frac{(i-p)^2+(j-q)^2}{\sigma_s^2}\right)\right]C_{pq}^{rs}, \qquad \text{(A32)}$$

where $C_{pq}^{rs} = \left[c_{pq}^{rs}\right]^+$ is rectified input from the complex cell in (A28). The on-center feedback from the bipole stage is provided by one-to-one projections:

$$C_{7ij}^{rs} = k_{ex}Z_{ij}^{rs}, \qquad \text{(A33)}$$

where $Z_{ij}^{rs} = \left[z_{ij}^{rs}\right]^+$ is the bipole output from the corresponding orientation *r* and scale *s* (equation A38). The off-surround feedback from the bipole stage is given by:

$$E_{7ij}^{s} = k_{inh}\sum_{r}\sum_{pq}\frac{E}{2\pi\sigma_s^2}\exp\left[-0.5\left(\frac{(i-p)^2+(j-q)^2}{\sigma_s^2}\right)\right]Z_{pq}^{rs}. \qquad \text{(A34)}$$

Summation of inhibitory feedback over all orientations *r* provides a spatial competition property, and suppresses hypercomplex cell activities, y_{ij}^{rs}, in the case of ambiguous boundary signals. This feedback is the part of competitive mechanism that breaks the stems of T-junctions from their

tops during figure-ground separation. Parameters are: A_3 =50, C=E=1; $\sigma_c = 0.5$, $\sigma_s = 1$, k_{ex} = 0.02, and k_{inh} = 0.2.

The orientation competition activities, n_{ij}^{rs}, receive rectified inputs $Y_{ij}^{rs} = \left[y_{ij}^{rs}\right]^+$ from the spatial competition stage (A30):

$$\frac{dn_{ij}^{rs}}{dt} = A_4\left(-n_{ij}^{rs} + (1-n_{ij}^{rs})\sum_{d\in\{1,\ldots,4\}} C^{dr}Y_{ij}^{ds} - (n_{ij}^{rs}+1)\sum_{d\in\{1,\ldots,4\}} E^{dr}Y_{ij}^{ds}\right). \quad \text{(A35)}$$

In (A35), orientation signals at each point (*i,j*) compete across orientations via term $E^{dr}Y_{ij}^{ds}$. Both the excitatory, C^{dr}, and inhibitory, E^{dr}, kernels have Gaussian profiles across orientations so that orthogonal orientation suppression is the strongest. The excitatory orientation competition kernel is:

$$C^{dr} = \frac{c}{2\pi\sigma_c^{\,2}}\exp\left[-0.5\left(\frac{d-r}{\sigma_c}\right)^2\right]. \quad \text{(A36)}$$

The inhibitory orientation competition kernel is:

$$E^{dr} = \frac{s}{2\pi\sigma_s^{\,2}}\exp\left[-0.5\left(\frac{d-r}{\sigma_s}\right)^2\right]. \quad \text{(A37)}$$

Parameters are: A_4 = 100, $\sigma_c = 0.5$, $\sigma_s = 0.75$, c = 0.5, and s = 0.75.

Level 5. Long-range bipole grouping. Bipole cell activities, z_{ij}^{rs}, cooperatively group hypercomplex cell inputs aligned in the same orientation as the bipole orientation preference, and allow boundaries to complete over gaps:

$$\frac{dz_{ij}^{rs}}{dt} = A_5\left(-z_{ij}^{rs} + H_{ij}^{rs} - (1+z_{ij}^{rs})H_{ij}^{Rs}\right), \quad R \perp r \ . \quad \text{(A38)}$$

The bipole input is defined by:

$$H_{ij}^{rs} = \left\lfloor g(A_{ij}^{rs}) + g(B_{ij}^{rs}) + k_N N_{ij}^{rs} - \theta_1 \right\rfloor, \quad \text{(A39)}$$

The two terms A_{ij}^{rs} and B_{ij}^{rs} sum collinear hypercomplex cell signals, N_{ij}^{rs}, from opposite sides of the bipole cell, where $N_{ij}^{rs} = \left[n_{ij}^{rs}\right]^+$ is the rectified output of the orientation competition stage (A35). The signal function:

$$g(x) = \frac{\lfloor x-\theta_2 \rfloor}{D_2 + \lfloor x-\theta_2 \rfloor} \quad \text{(A40)}$$

and threshold θ_1 are chosen so that both branches A_{ij}^{rs} and B_{ij}^{rs} must be sufficiently active to fire H_{ij}^{rs} in the absence of the direct bottom-up input N_{ij}^{rs} .The simplified bipole kernel includes only spatial pooling across the same orientation from both bipole branches:

$$A_{ij}^{rs} = \sum_{pq}\exp\left(-0.5\left(\frac{(i-p)^2}{\sigma_x^2} + \frac{(j-q-h)^2}{\sigma_y^2}\right)\right)N_{pq}^{rs}, \quad \text{(A41)}$$

and

$$B_{ij}^{rs} = \sum_{pq} \exp\left(-0.5\left(\frac{(i-p)^2}{\sigma_x^2} + \frac{(j-q+h)^2}{\sigma_y^2} \right) \right) N_{pq}^{rs} , \qquad \text{(A42)}$$

where $\sigma_x = 1$, $\sigma_y = 2$, and offset h = 5. These kernels are for the vertically oriented bipole. Each kernel A (top branch) and B (bottom branch) is elongated along the orientation it is pooling to facilitate grouping of the corresponding orientation across boundary gaps. Other orientation kernels are derived by rotation. Parameters are $D = 0.2$, $\theta_1 = 0.5$, $\theta_2 = 0.3$, $k_N = 1.5$. The last term in (A38) introduces competition between orthogonal orientations. Parameter $A_5 = 50$.

In a 2D image, 3D information such as occlusion is often represented by T-junctions. FACADE has proposed how T-junction detection and figure-ground separation occur without using explicit T-junction detectors. Simulation of the complete dynamics of boundary separation in depth would require large-scale simulations (Kelly and Grossberg, 2000; Grossberg and Yazdanbaksh, 2005), and were not implemented. Instead, it was assumed that after one chopstick boundary wins in the near depth, D1, these already demonstrated FACADE interactions will complete the boundary separation in depth. Algorithmically, a certain boundary orientation *r* won when the ratio, α, of the total boundary activation with orientation r to the total competing boundary *R* activation exceeds a threshold *T*:

$$\alpha = \frac{\sum_{ij} Z_{ij}^{rs}}{\sum_{ij} Z_{ij}^{Rs}} \geq T . \qquad \text{(A43)}$$

This ratio was computed in the circular neighborhood of a given junction. The radius of the circle was 5 cells. For a T-junction, the ratio of perpendicular orientations *r* and *R* in the neighborhood of T-junction was computed. For a Y-junction, the ratio of competing orientations, *r* and *R*, 45 ^{o} apart was computed. After the criterion (A43) was met, further V2 processing of the V1 bipole signal z_{ij}^{rs} (A38) was simplified by representing the corresponding boundaries by idealized 1-cell wide boundary activities, $\tilde{z}_{ij}^{rs}$, corresponding to the demonstrated ability of bipole cells to form sharp boundaries. These boundaries were positioned at the same locations as non-zero input boundary values I_{ij}. When one boundary wins in the near depth, it suppresses the same orientation via V2-to-V1 feedback in the far depth at the complex cell stage level (A28), thus resulting in a given boundary being represented only in one depth. The V2-to-MT projection in (A14) was calculated as the sum of the bipole activations across all orientations $Z_{ij}^{s} = \sum_{r} Z_{ij}^{rs}$, and then simplified with a 1-cell wide depth-separated boundary $\tilde{z}_{ij}^{s}$.

A summary of parameter choices for simulations and comparisons with parameters used in previously published models appear in Tables A1 and A2 for the motion and form streams, respectively.

Table A1
Comparison of present and previously published parameters for motion levels

	3D FORMOTION value	Equation Number	Grossberg et al., 2001 value
Level 2			
A_1	10	5	10
B_1	3	5	1
C_1	1	5	2
A_2	0.01	6	0.03
K_2	20	6	100
A_3	5	7	1
B_3	2	7	1
C_3	0.5	7	1
K_3	20	7	10
A_4	30	8	10
B_4	1	8	1
C_4	0.5	8	1
K_4	20	8	10
Level 3			
A_5	50	9	4
σ_x^s s=1	0.5	10	0.5
σ_y^s, s=1	1.5	10	1.5
σ_x^s, s=2	0.5	10	0.5
σ_y^s, s=2	2.5	10	2.5
θ_1, s=1	0.04	9	0.25
θ_2, s=2	0.08	9	0.5
Level 4			
A_6	50	11	20
C_6	5	11	10
D_6	100	11	no equiv
σ_x J	0.5	12	0.5
σ_y J	2.5	12	2.5
σ_x K	4	13	4
σ_y K	4	13	4
Level 5			
A_7	100	14	no equiv
K_e	1	14	no equiv
K_z	10	14	no equiv

K_b	0.12	14	no equiv
σ	6	15	no equiv
MT			
A_8	200	16	1
D_8	5	16	1
σ_x L	4	18	20
σ_y L	10	18	20
θ L	0.03	17	0
MST			
A_9	400	21	1
C_9	1	21	0.01
D_9	10	21	1

Table A2
Comparison of present and previously published parameters for form levels

Form System	3D Formotion value	Equation Number	Grossberg and Kelly, 2000 value
Simple cells			
σ_x, s=1	0.75	25, 26	0.75
σ_y, s=1	1	25, 26	1
σ_x, s=2	0.75	25, 26	1
σ_y, s=2	2.5	25, 26	1.25
Complex cell			
A_2	2	28	0.01
B_2	1	28	15
k_{ex}	25	28	no equiv
θ_{MT}	0.15	29	no equiv
D_2	10	28	1
Hypercomplex cell			
A_3	50	30	0.1
k_{ex}	0.02	33	1
k_{inh}	0.2	34	1
σ_c	0.5	31	1
σ_{surr}	1	32	2
Orientation comp			
A_4	100	35	1
σ_c	0.5	36	0.5
σ_{surr}	0.75	37	0.75
Bipoles			
A_5	50	38	1
D_2	0.2	40	0.1
k_N	1.5	39	0
θ_1	0.5	39	0.1
θ_2	0.3	40	0

References

Abbott, L.F., Sen, K., Varela, J.A., and Nelson, S.B. (1997). Synaptic depression and cortical gain control. *Science*, 275, 220-222.

Albright, T.D. (1984). Direction and orientation selectivity of neurons in visual area MT of the macaque. *Journal of Neurophysiology*, 52(6), 1106-30.

Anderson, B. L. and Barth, H. C. (2000). Motion-based mechanisms of illusory contour synthesis, *Neuron*, 24, 433-441.

Anderson, J.C. and Martin, K.A. (2002). Connection from cortical area V2 to MT in macaque monkey. *Journal of Comparative Neurology*, 443(1), 56-70.

Anderson, J.C., Binzegger, T., Martin, K.A. and Rockland., K.S. (1998). The connection from cortical area V1 to V5: a light and electron microscopic study. *Journal of Neuroscience*, 18(24):10525-40.

Anstis, S. M. (1990). Imperceptible intersections: the chopstick illusion. *In AI and the Eye, NY,Wiley*.

Bair, W., Cavanaugh, J.R., Smith, M.A. and Movshon, J.A. (2002). The timing of response onset and offset in macaque visual neurons. *Journal of Neuroscience*, 22(8), 3189-205.

Bakin, J.S., Nakayama, K., and Gilbert, C.D. (2000). Visual responses in monkey areas V1 and V2 to three-dimensional surface configurations. *Journal of Neuroscience*, 20(21), 8188-98.

Baloch, A.A. and Grossberg, S. (1997). A neural model of high-level motion processing: line motion and formotion dynamics. *Vision Research,* 37(21), 3037-59.

Baloch, A.A., Grossberg, S., Mingolla, E., and Nogueira, C.A.M. (1998). A neural model of first-order and second-order motion perception and magnocellular dynamics. *Journal of the Optical Society of America A*, 16, 953-978.

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

Berzhanskaya, J., Grossberg, S., and Mingolla, E. (2004). Motion-to-Form cortical projections and distortion of the position maps. *Vision Sciences Society*, 149, Sarasota, Florida.

Bowns, L. (1996). Evidence for a feature-tracking explanation of why type II plaids move in the vector sum direction at short durations. *Vision Research*, 36(22), 3685-3694.

Born, R.T. (2000). Center-surround interactions in the middle temporal visual area of the owl monkey. *Journal of Neurophysiology*. 84(5):2658-69.

Braddick O. (1993) Segmentation versus integration in visual motion processing. *Trends in Neurosciences,* 16(7):263-8.

Bradley, D.C. and Andersen, R.A. (1998) Center-surround antagonism based on disparity in primate area MT. *Journal of Neuroscience,* 18(18), 552-65.

Bradley, D.C., Chang G.C., and Andersen, R.A. (1998). Encoding of three-dimensional structure-from-motion by primate area MT neurons. *Nature*, 392(6677), 714-7.

Bregman, A.L. (1981). Asking the "what for" question in auditory perception. In M. Kubovy and J. R. Pomerantz. *Perceptual organization*. Hillsdale, NJ: Earlbaum Associate.

Bullier, J. (2001). Integrated model of visual processing. *Brain Research - Brain Research Reviews,* 36(2-3), 96-107.

Callaway, E.M. (1998). Local circuits in primary visual cortex of the macaque monkey. *Annual Review of Neuroscience*, 21, 47-74.

Cao, Y. and Grossberg, S. (2005). A laminar cortical model of stereopsis and 3D surface perception: Closure and da Vinci stereopsis. *Spatial Vision*, 18, 515-578.

Carandini, M. and Ferster, D. (1997). Visual adaptation hyperpolarizes cells of the cat striate

cortex. *Science*, 276, 949.
Carpenter, G.A. and Grossberg, S. (1991). *Pattern recognition by self-organizing neural networks*. Cambridge, MA: MIT Press.
Chance, F.S., Nelson., S.B., and Abbott, L.F. (1998). Synaptic depression and the temporal response characteristics of V1 cells. *Journal of Neuroscience*, 18(12), 4785-4799.
Chey, J., Grossberg, S., and Mingolla, E. (1997). Neural dynamics of motion processing and speed discrimination. *Vision Research,* 38, 2769-2786.
Chey, J., Grossberg, S., and Mingolla, E. (1998). Neural dynamics of motion processing and speed discrimination. *Vision Research*, 38(18), 2769-86.
Cox, M.J. and Derrington, A.M. (1994). The analysis of motion of two-dimensional patterns: do Fourier components provide the first stage? *Vision research*, 34(1), 59-72.
DeAngelis, G.C. and Uka, T. (2003). Coding of horizontal disparity and velocity by MT neurons in the alert macaque. *Journal of Neurophysiology*, 89(2):1094-111.
DeYoe, E.A. and Van Essen, D.C. (1985). Segregation of efferent connections and receptive field properties in visual area V2 of the macaque. *Nature*, 317(6032):58-61.
Duncan, R.O., Albright, T.D., and Stoner, G.R. (2000). Occlusion and the interpretation of visual motion: perceptual and neuronal effects of context. *Journal of Neuroscience*, 20(15), 5885-97.
Duncker, K. (1937). Induced motion. In W.E. Ellis (Ed.), *A Sourcebook of Gestalt Psychology*. London: Routledge and Kegan Paul. (Original work published in 1929).
Enroth-Cugell, C. and Robson, J. (1966). The contrast sensitivity of retinal ganglion cells of the cat. *Journal of Physiology (London),* 187, 517-552.
Fang, L. and Grossberg, S. (2004). How are the surface lightnesses of complex stereograms assigned to the correct depths? *Vision Sciences Society*, 152, Sarasota, Florida.
Ferrera, V.P. and Wilson, H.R. (1987). Direction specific masking and the analysis of motion in two dimensions. *Vision Research.* 27(10), 1783-96.
Francis, G. and Grossberg, S. (1996). Cortical dynamics of form and motion integration: persistence, apparent motion and illusory contours. *Vision Research*, 36(1), 149-73.
Fried, S.I., Münch, T.A. and Werblin, F.S. (2002). Mechanisms and circuitry underlying directional selectivity in the retina. *Nature*, 420, 411-414.
Fu, Y.X., Shen, Y., Gao, H., and Dan, Y. (2004) Asymmetry in visual cortical circuits underlying motion-induced perceptual mislocalization. *Journal of Neuroscience,* 24(9), 2165-71.
Grunewald, A., Bradley, D.C. and Andersen, R.A. (2002). Neural correlates of structure-from-motion perception in macaque V1 and MT. *Journal of Neuroscience*, 22(14):6195-207.
Grossberg, S. (1973). Contour enhancement, short-term memory, and constancies in reverberating neural networks. *Studies in Applied Math*, 52, 213-257.
Grossberg, S. (1978). A theory of human memory: Self-organization and performance of sensory-motor codes, maps, and plans. In R. Rosen and F. Snell (Eds.), *Progress in theoretical biology*, Volume 5. New York: Academic Press, pp. 233-374.
Grossberg, S. (1980). How does a brain build a cognitive code? *Psychological Review*, 87(1), 1-51.
Grossberg, S. (1988). Nonlinear neural networks: principles, mechanisms, and architectures. *Neural Networks*, 1, 12-61.
Grossberg, S. (1991). Why do parallel cortical systems exist for the perception of static form and moving form? *Perception and Psychophysics*, 49, 117-141.

Grossberg, S. (1994). 3-D vision and figure-ground separation by visual cortex. *Perception & Psychophysics*, 55, 48-121.

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

Grossberg, S. (1999). How does the cerebral cortex work? Learning, attention and grouping by the laminar circuits of visual cortex. *Spatial Vision,* 12**,** 163-185.

Grossberg, S. (2000). The complementary brain: Unifying brain dynamics and modularity. *TICS* 4(6), 233-46.

Grossberg, S. and Howe, P.D. (2003). A laminar cortical model of stereopsis and three-dimensional surface perception. *Vision Research*, 43(7), 801-29.

Grossberg, S. and Mingolla, E. Ross (1997). Visual brain and visual perception: How does the cortex do perceptual grouping? *Trends in Neurosciences,* 20, 106-111.

Grossberg, S., Mingolla, E., and Viswanathan, L. (2001). Neural dynamics of motion integration and segmentation within and across apertures. *Vision Research*, 41, 2351-2553.

Grossberg, S. and Pearson, L. (2006). Laminar cortical dynamics of cognitive and motor working memory, sequence learning and performance: Towards a unified theory of how the cerebral cortex works. CAS/CNS Tech Report-2006-002. Submitted for publication.

Grossberg, S. and Raizada, R.D. (2000). Contrast-sensitive perceptual grouping and object-based attention in the laminar circuits of primary visual cortex. *Vision Research*, 40(10-12),1413-32.

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

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

Grossberg, S. and Seitz, A. (2003). Laminar development of receptive fields, maps and columns in visual cortex: the coordinating role of the subplate. *Cerebral Cortex*, 13(8), 852-63.

Grossberg, S. and Swaminathan, G. (2004). A laminar cortical model for 3D perception of slanted and curved surfaces and of 2D images: development, attention, and bistability. *Vision Research*, 44(11), 1147-1187.

Grossberg S. and Williamson J.R. (2001). A neural model of how horizontal and interlaminar connections of visual cortex develop into adult circuits that carry out perceptual grouping and learning. *Cerebral Cortex*, 11(1), 37-58.

Grossberg, S. and Yazdanbaksh, A. (2005). Laminar cortical dynamics of 3D surface perception: Stratification, transparency, and neon color spreading. *Vision Research*, 45, 1725-1743.

Grzywacz, N. and Yuille, A. (1991). Theories for the visual perception of local velocity and coherent motion. In *Computational models of visual processing* (Cambridge, MA, MIT press).

Hildreth, E.C. (1983). *The measurement of visual motion*. (Cambridge, MA, MIT press)

Hochstein, S. and Shapley, R.M. (1976a). Linear and nonlinear spatial subunits in Y cat retinal ganglion cells. *Journal of Physiology*, 262(2), 265-84.

Hochstein, S. and Shapley R.M. (1976b). Quantitative analysis of retinal ganglion cell classifications. *Journal of Physiology*, 262(2), 237-64.

Hupé, J.M., James, A.C., Payne, B.R., Lomber, S.G., Girard, P. and Bullier, J. (1998). Cortical feedback improves discrimination between figure and background by V1, V2 and V3 neurons. *Nature*, 394(6695):784-7.

Hupé, J.M. and Rubin, N. (2003). The dynamics of bi-stable alternation in ambiguous motion displays: a fresh look at plaids. Vision Research, 43(5):531-48.
Jones, H.E., Grieve, K.L., Wang, W., and Sillito, A.M. (2001). Surround suppression in primate V1. *Journal of Neurophysiology.* 86, 2011-28.
Kanizsa, G. (1979). *Organization in vision: essays on Gestalt perception*. NewYork: Praeger Press.
Kelly, F. J. and Grossberg, S. (2000). Neural dynamics of 3-D surface perception: Figure-ground separation and lightness perception. *Perception & Psychophysics*, 62, 1596-1619.
Kim, J. and Wilson, H.R. (1993). Dependence of plaid motion coherence on component grating directions. *Vision Research,* 33(17), 2479-89.
Kennedy, H. and Bullier, J. (1985). A double-labeling investigation of the afferent connectivity to cortical areas V1 and V2 of the macaque monkey. *Journal of Neuroscience*, 5, 2815-30.
Lidén, L. and Mingolla, E. (1998) Monocular occlusion cues alter the influence of terminator motion in the barber pole phenomenon. *Vision Research,* 38(24):3883-98.
Liden, L. and Pack, C. (1999). The role of terminators and occlusion cues in motion integration and segmentation: A neural network model. *Vision Research,* 39(19), 3301-20.
Livingstone, M.S. (1998). Mechanisms of direction selectivity in macaque V1. *Neuron,* 20(3):509-26.
Livingstone, M.S., Conway, B.R. (2003). Substructure of direction-selective receptive fields in macaque V1. *Journal of Neurophysiology,* 89(5):2743-59.
Livingstone, M.S. and Hubel, D.H. (1984). Anatomy and physiology of a color system in the primate visual cortex. *Journal of Neuroscience*. 4 (1), 309-356.
Lorenceau, J. and Alais, D. (2001). Form constraints in motion binding. *Nature Neuroscience*, 4, 745-751.
Malach, R., Schriman, T.D., Harel, M., Tootell, R.B.H. and Malonek, D. (1997) Organization of intrinsic connections of owl monkey MT. *Cerebral cortex*, 7(94), 386-93.
McGraw, P.V., Walsh V., Barrett B.T. (2004) Motion-sensitive neurones in V5/MT modulate perceived spatial position. *Current Biology*, 14(12):1090-1093.
Mishkin, M., Ungerleider, L. G., and Macko, K. A. (1983). Object vision and spatial vision: Two cortical pathways. *Trends in Neurosciences,* 6, 414-417.
Movshon, J.A. and Newsome, W.T. (1996). Visual response properties of striate cortical neurons projecting to area MT in macaque monkeys. *Journal of Neuroscience*. 16(23), 7733-41.
Muller, J.R., Metha, A.B., Krauskopf, J. and Lennie, P. (2001) Information conveyed by onset transients in responses of striate cortical neurons. *Journal of Neuroscience*. 21(17) 6987-6990.
Murakami, I. (1999). Motion-transparent inducers have different effects on induced motion and motion capture. *Vision Research*, 39(9), 1671-81.
Murthy, A., Humphrey, A.L. (1999) Inhibitory contributions to spatiotemporal receptive-field structure and direction selectivity in simple cells of cat area 17. *Journal of Neurophysiology*, 81(3):1212-24.
Nakayama, K., Shimojo, S., and Silverman, G.H. (1989). Stereoscopic depth: its relation to image segmentation, grouping, and the recognition of occluded objects. *Perception,* 18(1), 55-68.
Nijhawan, R. (1994). Motion extrapolation in catching. *Nature*, 370(6487), 256-7.
Pack, C.C., Berezovskii, V.K., and Born, R.T. (2001). Dynamic properties of neurons in cortical area MT in alert and anaesthetized macaque monkeys. *Nature*, 414(6866), 905-8.

Pack, C.C. and Born, R. T. (2001) Temporal dynamics of a neural solution to the aperture problem in visual area MT of macaque brain. *Nature*, 409(6823):1040-2.

Pack, C.C., Gartland, A.J., and Born, R.T. (2004). Integration of contour and terminator signals in visual area MT of alert macaque. *Journal of Neuroscience,* 24(13), 3268-80.

Palanca, B.J. and DeAngelis, G.C. (2003). Macaque middle temporal neurons signal depth in the absence of motion. *Journal of Neuroscience*, 23, 7647-7658.

Palmer, E. M. and Kellman, P. J. (2001). The Aperture-capture effect: Misperceived forms in dynamic occlusion displays. *Journal of Vision,* 1(3), 381a.

Ponce, C.R., Lomber, S.G., and Born R.T. (2006) Modality-specific changes in tuning properties of MT neurons during inactivation of the indirect (V2/V3) pathway in alert macaques. *Society for Neuroscience abstracts*, Atlanta, GA.

Raizada RD. and Grossberg S. (2003). Towards a theory of the laminar architecture of cerebral cortex: Computational clues from the visual system. *Cerebral Cortex,* 13(1), 100-13.

Reichardt W. in *Sensory Communication*, W.A. Rosenblith, ed. (Wiley, New York, 1961), p. 303.

Reynolds, J., Chelazzi, L. and Desimone, R. (1999). Competitive mechanisms subserve attention in macaque areas V2 and V4. *Journal of Neuroscience*, 19, 1736-53.

Rockland, K.S. (2002). Visual cortical organization at the single axon level: a beginning. *Neuroscience Research*, 42(3):155-66.

Rockland, K.S. (1995). Morphology of individual axons projecting from area V2 to MT in the macaque. *Journal of Comparative Neurology*, 355(1), 15-26.

Rockland, K.S. (1992) Laminar distribution of neurons projecting from area V1 to V2 in macaque and squirrel monkeys. *Cerebral Cortex,* 2(1):38-47.

Rockland, K.S and Pandya, D.N. (1981) Cortical connections of the occipital lobe ian the rhesus monkey: interconnections between areas 17, 18, 19 and the superior temporal sulcus. *Brain Research,* 212(2):249-70.

Rust, N.C., Majaj, N.J., Simoncelli, E.P. and Movshon, J.A. (2002). Gain control in Macaque area MT is directionally selective. *SFN 2002*, Orlando, Florida.

Schiller, P.H. and Logothetis, N.K. (1990). The color-opponent and broad-band channels of the primate visual system. *Trends in Neurosciences,* 13 (10), 392-398.

Schmidt, K.E., Goebel, R., Lowel, S. and Singer, W. (1997) The perceptual grouping criterion of colinearity is reflected by anisotropies of connections in the primary visual cortex. *Eur J Neurosci.* 9(5), 1083-9

Shimojo, S., Silverman, G.H. and Nakayama, K. (1989). Occlusion and the solution to the aperture problem for motion. *Vision Research*, 29(5), 619-26.

Shipp, S. and Zeki, S. (1989) The organization of connections between areas V5 and V1 in Macaque visual cortex. European Journal of Neuroscience, 1, 309-332.

Shipp, S. and Zeki, S. (1985) Segregation of pathways leading from area V2 to areas V4 and V5 of macaque monkey visual cortex. *Nature,* 315(6017):322-5.

Silvanto J., Lavie N., Walsh V. (2005) Double dissociation of V1 and V5/MT activity in visual awareness. *Cerebral Cortex*, 15(11):1736-41.

Sincich LC and Horton JC. (2003). Independent projection streams from macaque striate cortex to the second visual area and middle temporal area. *Journal of Neuroscience,* 23(13):5684-92.

Sincich, L.C. and Horton, J.C. (2002) Divided by cytochrome oxidase: a map of the projections from V1 to V2 in macaques. *Science,* 295(5560):1734-7.

Sincich, L.C. and Blasdel G.G. (2001). Oriented axon projections in primary visual cortex of the monkey. *Journal of Neuroscience*, 21, 4416-4426.

Snowden, R.J., Treue, S., Erickson RG., and Andersen RA. (1991). The response of area MT and V1 neurons to transparent motion. *Journal of Neuroscience,* 11(9), 2768-85.

Stoner, G.R., Albright, T.D., and Ramachandran, V.S. (1990). Transparency and coherence in human motion perception. *Nature,* 344(6262), 153-5.

Stoner, G.R. and Albright, T.D. (1996). The interpretation of visual motion: Evidence for surface segmentation mechanisms. *Vision Research*, 36(9), 1291-310.

Stoner, G.R. and Albright, T.D. (1998). Luminance contrast affects motion coherency in plaid patterns by acting as a depth-from-occlusion cue. *Vision Research*, 38(3), 387-401.

Treue, S. and Maunsell, J.H. (1999). Effects of attention on the processing of motion in macaque middle temporal and medial superior temporal visual cortical areas. *Journal of Neuroscience*, 19(17), 7591-602.

Trueswell, J.C. and Hayhoe, M.M. (1993). Surface segmentation mechanisms and motion perception. *Vision Research,* 33(3), 313-23.

Vallortigara, G., Bressan, P., and Bertamini, M. (1988). Perceptual alternations in stereokinesis. *Perception*, 17, 4.-31.

van Santen, J.P. and Sperling, G. (1985) Elaborated Reichardt detectors. *J Opt Soc Am A.,* 2(2),300-21.

Varela, J.A., Sen, K., Gibson, J., Fost, J., Abbott, L.F., and Nelson., S.B. (1997). A quantitative description of short-term plasticity at excitatory synapses in Layer 2/3 of rat primary visual cortex. *Journal of Neuroscience*, 17(20), 7926-7940.

Wallach, H. (1935). On the visually perceived direction of motion. *Psychologische Forschung*, 20, 325-380.

Weiss, Y. and Adelson, E. H. (2000). Adventures with gelatinous ellipses—constraints on models of human motion analysis. *Perception*, 29, 543-566.

Weiss, Y., Simoncelli, E.P., and Adelson, E.H. (2002). Motion illusions as optimal percepts. *Nature Neuroscience,* 5(6), 598-604.

Wilson, H.R., Ferrera, V.P., and Yo, C. (1992). A psychophysically motivated model for two-dimensional motion perception. *Visual Neuroscience*, 9(1), 79-97.

Whitney, D. and Cavanagh, P. (2000). Motion distorts visual space: shifting the perceived position of remote stationary objects. *Nature Neuroscience*, 3(9), 954-9.

Wuerger, S., Shapley, R., and Rubin, N. (1996). 'On the visually perceived direction of motion' by Hans Wallach: 60 years later. *Perception*, 25, 1317-1367.

Xiao, D.K., Raiguel, S., Marcar, V. and Orban, G.A. (1997). The spatial distribution of the antagonistic surround of MT/V5 neurons. *Cerebral Cortex,* 7(7), 662-77.

Xu, X., Bonds, A.B., and Casagrande, VA. (2002). Modeling receptive-field structure of koniocellular, magnocellular, and parvocellular LGN cells in the owl monkey (Aotus trivigatus). *Visual Neuroscience*, 19(6), 703-11.

Yabuta, N.H., Sawatari, A. and Callaway, E.M. (2001) Two functional channels from primary visual cortex to dorsal visual cortical areas. *Science*, 292(5515):297-300.

Yabuta, N.H. and Callaway, E.M. (1998) Functional streams and local connections of layer 4C neurons in primary visual cortex of the macaque monkey. *Journal of Neuroscience,* 18(22):9489-99.

Zhou, H., Friedman, H.S., and von der Heydt, R. (2000). Coding of border ownership in monkey visual cortex. *Journal of Neuroscience,* 20, 6594-6611.