A Laminar Cortical Model of Stereopsis and 3D Surface Perception: Closure and da Vinci Stereopsis

Yongqiang Cao and Stephen Grossberg[1]

Department of Cognitive and Neural Systems
and
Center for Adaptive Systems
Boston University
677 Beacon Street, Boston, MA 02215, USA

Running title: Laminar cortical model of depth perception

September 2004
Technical Report CAS/CNS-TR-2004-007
All correspondence should be addressed to
Professor Stephen Grossberg
Department of Cognitive and Neural Systems
Boston University
677 Beacon Street
Boston, MA 02215
Phone: 617-353-7858
Fax: 617-353-7755
Email:steve@bu.edu

[1]Authorship in alphabetical order. Y.C. was supported by the Natural Sciences and Engineering Research Council of Canada and the National Science Foundation (NSF SBE-0354378). SG was supported by the Air Force Office of Scientific Research (AFOSR F49620-01-1-0397), the National Science Foundation (NSF SBE-0354378), and the Office of Naval Research (ONR N00014-01-1-0624).

Abstract
A laminar cortical model of stereopsis and 3D surface perception is developed and simulated. The model describes how monocular and binocular oriented filtering interact with later stages of 3D boundary formation and surface filling-in in the LGN and cortical areas V1, V2, and V4. It proposes how interactions between layers 4, 3B, and 2/3 in V1 and V2 contribute to stereopsis, and how binocular and monocular information combine to form 3D boundary and surface representations. The model includes two main new developments: (1) It clarifies how surface-to-boundary feedback from V2 thin stripes to pale stripes helps to explain data about stereopsis. This feedback has previously been used to explain data about 3D figure-ground perception. (2) It proposes that the binocular false match problem is subsumed under the Gestalt grouping problem. In particular, the disparity filter, which helps to solve the correspondence problem by eliminating false matches, is realized using inhibitory interneurons as part of the perceptual grouping process by horizontal connections in layer 2/3 of cortical area V2. The enhanced model explains all the psychophysical data previously simulated by Grossberg and Howe (2003), such as contrast variations of dichoptic masking and the correspondence problem, the effect of interocular contrast differences on stereoacuity, Panum's limiting case, the Venetian blind illusion, stereopsis with polarity-reversed stereograms, and da Vinci stereopsis. It also explains psychophysical data about perceptual closure and variations of da Vinci stereopsis that previous models cannot yet explain.

Keywords: Visual cortical; Stereopsis; Depth perception; Binocular vision; Surface perception; Lightness perception; Monocular-binocular interactions; V1; V2; V4; LAMINART model

1. Introduction

Understanding how humans and other animals see the world in depth is an essential first step in understanding many visual behaviors. This article describes a model of how the LGN and cortical areas V1, V2 and V4 utilize both monocular and binocular visual information to produce three-dimensional (3D) surface percepts. The article builds upon the 3D LAMINART model of Grossberg and Howe (2003), which was used to explain and simulate a wide range of data about stereopsis and 3D planar surface perception. This model, shown in Figure 1, went beyond previous analyses of stereopsis in several ways. First, it provided a refined laminar model of stereopsis in V1 which clarifies the role of cells in cortical layers 4, 3B, and 2/3A. In so doing, the model revised how the disparity energy model (Ohzawa, DeAngelis and Freeman, 1990) achieves stereopsis, in a manner that is more consistent with recent data. Second, the model showed how monocular and binocular information may be combined and selected in V2 to form 3D boundary representations. Third, the model proposed how these 3D boundaries give rise to visible 3D surface percepts in V4. These model processes were used to explain and simulate a much larger set of neurophysiological, anatomical, and psychophysical data about stereopsis and 3D surface perception than had previously been possible. The 3D LAMINART model achieved these explanatory successes without including some processes that had been needed to explain other data about visual perception and neuroscience in previous modeling studies. These processes include perceptual grouping and surface-to-boundary feedback. Perceptual grouping was modeled in a laminar model of V1 and V2, called the LAMINART model, that did not incorporate binocular interactions (Grossberg, 1999a, 1999b; Grossberg, Mingolla and Ross, 1997; Grossberg and Raizada, 2000; Raizada and Grossberg, 2001; Grossberg and Williamson, 2001), but rather focused on explaining data about perceptual development, learning, grouping, and attention. It was noted in Grossberg and Howe (2003) how perceptual grouping might consistently be added to the 3D LAMINART model, but since this refinement was not needed to explain the targeted data, it was not further pursued. Grossberg and Swaminathan (2004) did make such a generalization in order to simulate data about 3D perception of slanted and curved 3D objects and 2D pictures of such objects, including a simulation of how a 2D image of a Necker cube causes rivalrous 3D boundary and surface representations of the cube. This advance required the introduction of neural circuits capable of 3D grouping. The Grossberg and Swaminathan study predicted how rules for 2D grouping using cells that obey a *bipole property* could be consistently generalized to carry out 3D groupings across depths, as well as at angles and smooth changes of curvature. The present version of the 3D LAMINART model also incorporates 3D bipole grouping laws, but refines them to clarify how 3D grouping helps to solve the classical correspondence problem.

The second innovation of the present work concerns the role of surface-to-boundary feedback in determining the 3D surfaces that are seen during stereopsis. Many experiments and modeling studies have supported the prediction of Grossberg and colleagues (e.g., Cohen and Grossberg, 1984; Grossberg, 1984; Grossberg and Mingolla, 1985a, 1985b) that visual perception is based on interacting percepts of boundaries and surfaces that obey complementary computational rules. This state of affairs raises the basic question: How can boundaries and surfaces interact to overcome their complementary deficiencies and generate a consistent percept? It has been proposed (e.g., Grossberg, 1994, 1997; Kelly and Grossberg, 2000) how feedback between cortical boundary and surface processing streams helps to transform boundary-surface complementarity into perceptual consistency and, in so doing, helps to separate figures from their backgrounds. Here we show how such surface-to-boundary feedback can help

to explain some basic stereopsis percepts, notably percepts that involve closure properties of figures and new variations of da Vinci stereopsis.

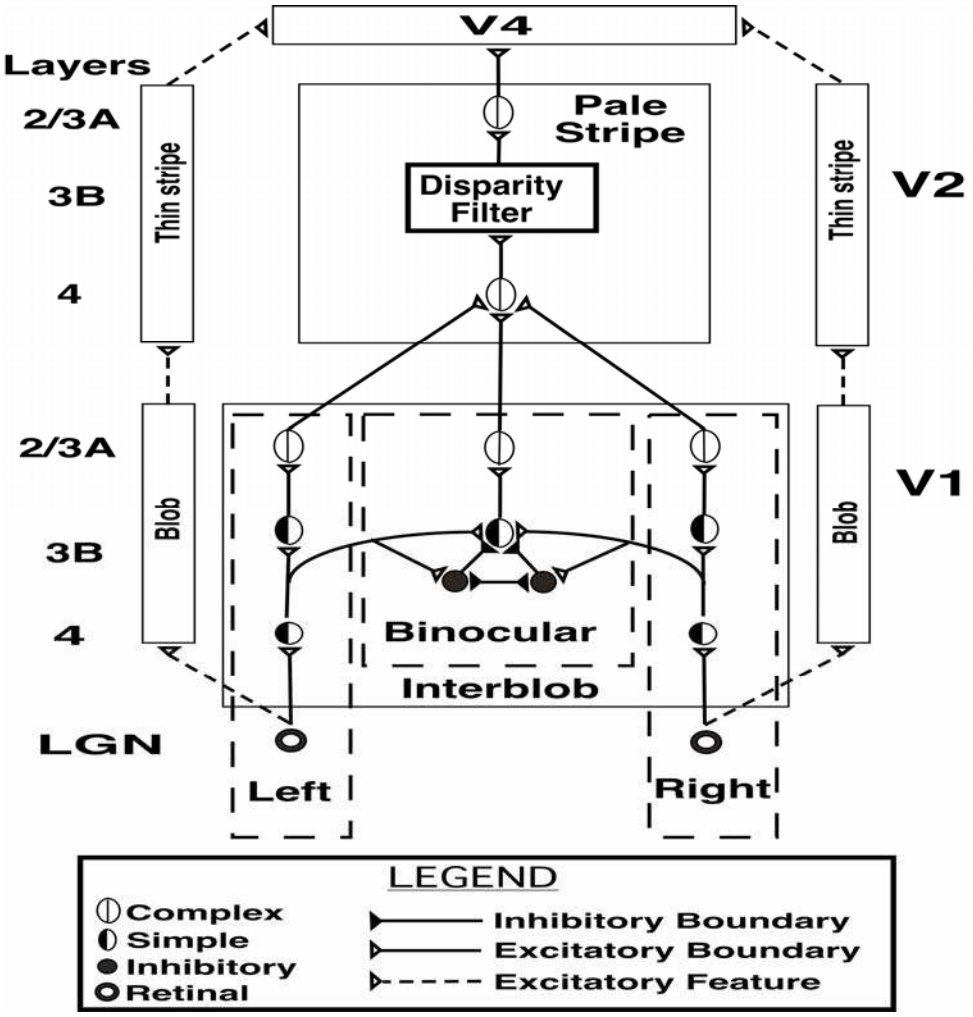


Figure 1. A model circuit diagram (Grossberg and Howe, 2003). [Reproduced with permission from Grossberg and Howe (2003).]

The model achieves these goals by embodying seven basic constraints in its neural circuitry. Constraints 1-4 and 6 were articulated in Grossberg and Howe (2003). Constraints 5 and 7 are needed to explain stereopsis data that go beyond the explanatory range of Grossberg and Howe (2003). Neurophysiological and anatomical data that support each of the hypotheses 1-4 and 6 are summarized in Grossberg and Howe (2003).

(1) *Reconciles contrast-specific binocular fusion with contrast-invariant boundary perception.* It is well known that only edges in the left and right retinal images that have the same contrast polarity (i.e., their luminance gradients have the same signs) can be binocularly fused to form a percept of depth (Howard and Rogers, 1995). Otherwise expressed, binocular fusion obeys the *same-sign hypothesis*. See Figure 2. However, fused boundaries must also be able to form around objects whose contrast polarity with respect to the background can reverse along their perimeters (Grossberg, 1994). In other words, binocular boundaries need to be

represented in a contrast-invariant way. How can the brain reconcile contrast-specific fusion with the need to form contrast-invariant object boundaries? The model proposes that both constraints are realized by interactions between cells in layers 4, 3B, and 2/3A of cortical area V1 interblob. See Figure 1.

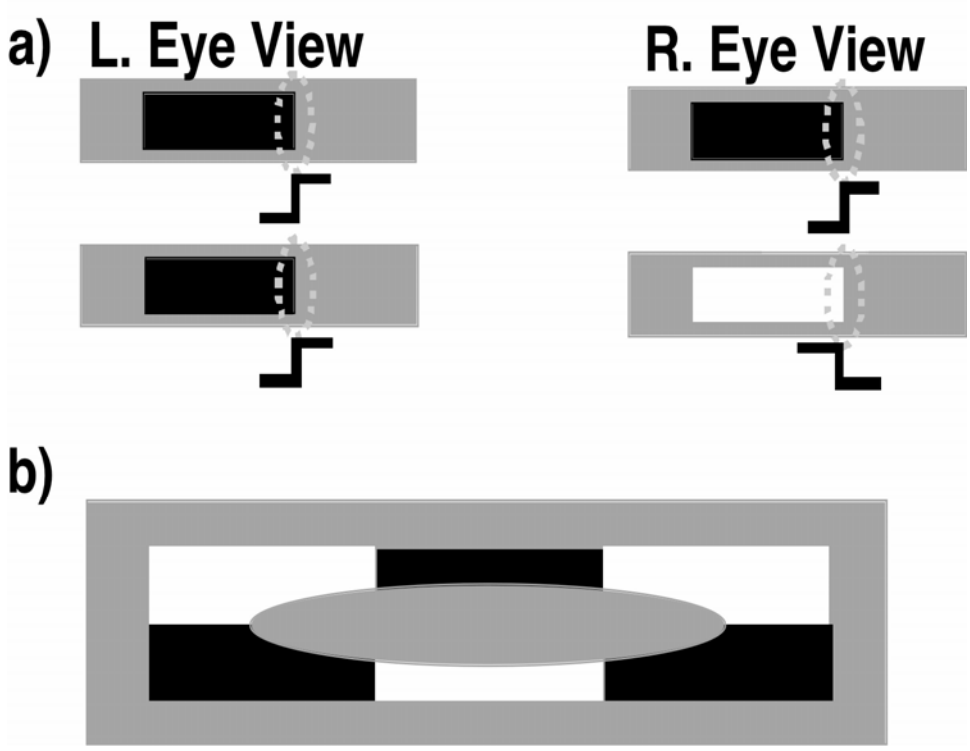


Figure 2. (a) The same-sign hypothesis: only edges that have the same contrast polarity can be stereoscopically fused to produce a percept of depth; (b) As it is traversed, the boundary of the ellipse changes its contrast polarity relative to the background, thereby illustrating the need for object boundaries to be represented in a contrast-invariant manner. [Reproduced with permission from Grossberg and Howe (2003).]

(2) *Implements the contrast magnitude constraint on binocular fusion.* The brain needs to determine which of the many potential edges in the two retinal images should be binocularly fused, since veridical stereoscopic depth perception will occur only if the two edges belong to the same object. This is commonly referred to as the *correspondence problem* (Howard and Rogers, 1995; Julesz, 1971). An early step in solving the correspondence problem is to binocularly fuse only edges with the same contrast polarity and approximately the same magnitude of contrast (McKee et al., 1994). This constraint naturally arises when the brain fuses edges that derive from the same objects in the world. The model satisfies this constraint through interactions between excitatory and inhibitory cells in layer 3B of V1 that endow the binocular cells there with an *obligate property* (Poggio, 1991), whereby they respond preferentially to left and right eye inputs of approximately equal contrast. See Figure 1.

(3) *Solves the correspondence problem.* Even if all binocular matches are of the same contrast polarity and similar contrast magnitude, there can still exist many false binocular matches between edges that did not derive from the same objects. See Figure 3.

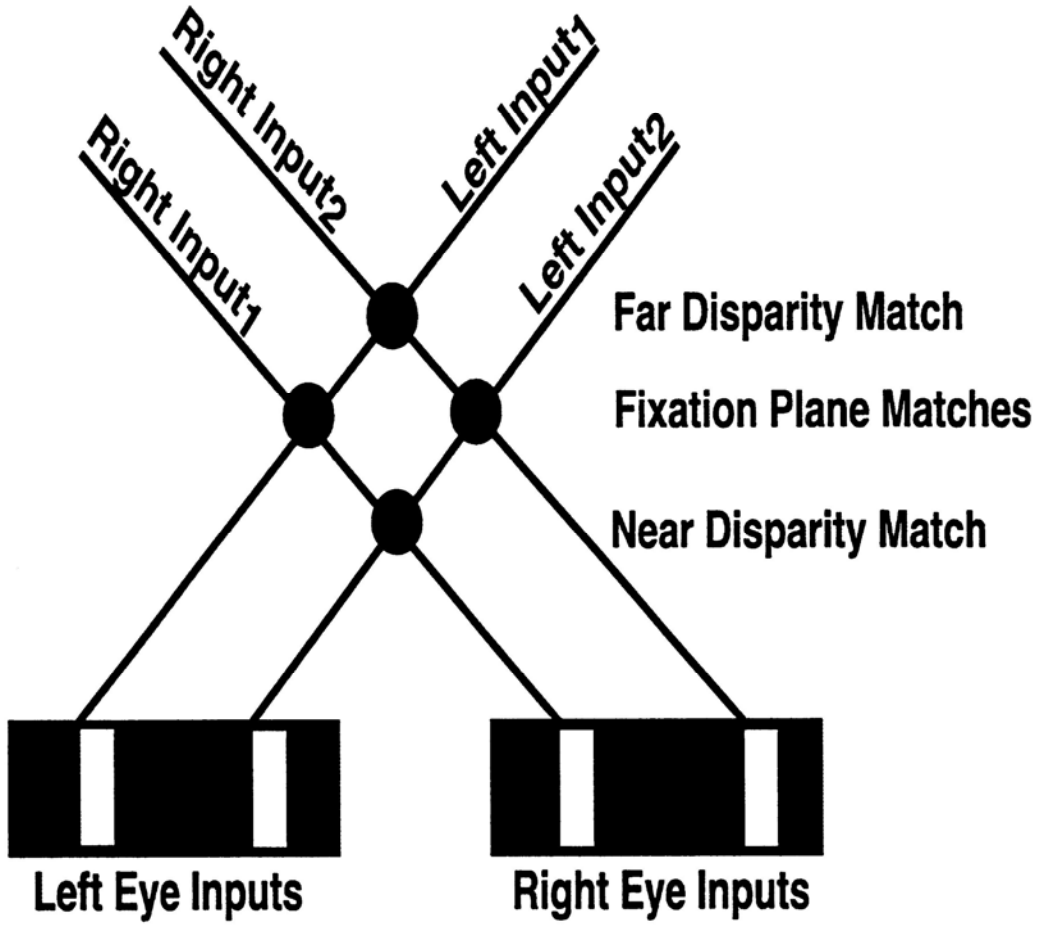


Figure 3. The V2 disparity filter. The V1 binocular boundaries network matches an edge in one retinal image with every other edge in the other retinal image whose relative disparity is not too great, that has the same contrast polarity and whose magnitude of contrast is not too different. In response to this image, the V1 boundary network creates four matches, with the two not in the fixation plane being false matches between edges that do not correspond to the same object. As described in the text, these false matches are suppressed by the disparity filter in V2, wherein each neuron is inhibited by every other neuron that shares either of its monocular inputs (i.e., shares a monocular line-of-sight represented by the solid lines; "line-of-sight inhibition"). Note in particular that the solid lines that represent the monocular lines-of-sight also represent the allelotropic shifts: an edge in the left retinal image is shifted to the right for matches increasingly further away whereas an edge in the right retinal image is shifted in the opposite direction.

This problem has often been approached by imposing a *unique-matching rule*, which states that any given feature in one retinal image is matched at most with one feature in the other retinal image (Grimson, 1981; Marr and Poggio, 1976). However, this rule fails in situations like Panum's limiting case (Gillam et al., 1995; McKee et al., 1995; Panum, 1858) where a bar presented to one eye is simultaneously matched to two separate bars presented to the other eye. The 3D LAMINART model does not enforce unique matches. Rather, the model encourages them by using a *disparity filter* (Grossberg and McLoughlin, 1997; McLoughlin and Grossberg, 1998). This disparity filter uses inhibition between active cells that represent different depths in

order to encourage unique matches. In Grossberg and Howe (2003), the disparity filter was located between layers 4 and 2/3 within the pale stripes of cortical area V2, and used a combination of line-of-sight inhibition and cyclopean inhibition, where cyclopean inhibition acts across depth between all cells at a given position. In the present study, the disparity filter is proposed to occur as part of the perceptual grouping process within layer 2/3 (see item (5) below). It was also found that the addition of surface-to-boundary feedback (see item (7) below), combined with the perceptual grouping process, made unnecessary the cyclopean inhibition that was also used in the Grossberg and Howe (2003) disparity filter. Thus, in the present model, line-of-sight inhibition helps to solve the correspondence problem as part of the perceptual grouping process.

(4) *Combines monocular and binocular information to form depth percepts.* Although Panum's limiting case may seem to be a laboratory curiosity, many naturally occurring situations lead to visual input where there is only one edge seen by one eye and two possible edges with which to match it seen by the other eye. For example, due to the lateral displacement of the eyes, an object's edge that is seen by one eye may be occluded in the other eye, as occurs during da Vinci stereopsis (Nakayama and Shimojo, 1990). Despite this lack of binocular information, the monocularly viewed region has a definite depth conferred to it by the binocularly viewed parts of the scene. The brain can thus utilize monocular information to build up seamless 3D percepts of the world. Indeed, in experiments involving Panum's limiting case, varying the relative contrast of the bars alters the perception of depth in a manner that reveals clear monocular-binocular interactions (Smallman and McKee, 1995). Dichoptic masking, where an object presented to one eye is obscured (i.e., masked) by one presented to the other eye, illustrates a third way in which monocular and binocular information may interact (McKee et al., 1994).

Once monocular information is included, the problem immediately arises of how to combine monocular and binocular boundaries. This is a problem because monocular boundaries do not have a definite depth associated with them. How, then, can we decide to which depth they should be assigned? A proposed approach to this *Monocular-Binocular Interface Problem* was suggested (Grossberg, 1994, 1997) in order to explain data about 3D figure-ground perception. The same hypothesis was shown by Grossberg and Howe (2003) to help explain many data about 3D surface perception; namely, the model assumes that the outputs of the monocular boundary cells are added to all depth planes in the pale stripes of cortical area V2 along their respective lines-of-sight, possibly in layer 4. The disparity filter, which helps to solve the correspondence problem, also solves the Monocular-Binocular Interface Problem by automatically eliminating most of the monocular boundaries that are not at the correct depths.

(5) *Forms 3D perceptual groupings that eliminate false matches*. Perceptual grouping in the model is carried out by pyramidal cells in layer 2/3 of the V2 pale stripes. Interactions between these cells realize a *bipole property*, whereby cells that are (approximately) colinear and coaxial with respect to one another across space can excite each other via long-range horizontal connections. These long-range horizontal connections also activate inhibitory interneurons that inhibit each other and nearby pyramidal cells via short-range disynaptic inhibition. This balance of excitation and inhibition at target cells helps to implement the bipole property: When the inducing stimulus (e.g., a pacman that defines a Kanizsa square) is only on one side, it excites the corresponding oriented receptive fields of layer 2/3A cells, which send out long-range horizontal excitation onto the target cell. However, this excitation also activates a commensurate amount of disynaptic inhibition. This creates a case of "one-against-one", and the target cell is not excited above-threshold. Suprathreshold cell activity can be modulated by input from long-

range horizontal connections on one side if the cell also receives bottom-up input (Bringuier, Chavane, Glaeser and Frégnac, 1999; Crook, Engelmann and Lowel, 2002). The combined bottom-up and horizontal input from one side can overcome the disynaptic inhibition from the inhibitory interneurons and thus can activate the cell. These modulations play an important role in the spreading of attention (Grossberg and Raizada, 2000; Ito and Gilbert, 1999; Roelfsema, Lamme and Spekreijse, 1998; Roelfsema and Spekreijse, 1999), the grouping of 2D and 3D planar percepts (Bakin et al., 2000; Kapadia, Ito, Gilbert and Westheimer, 1995; Polat, Mizobe, Pettet, Kasamatsu and Norcia, 1998), and the grouping of 3D slanted and curved percepts, as discussed in Grossberg and Swaminathan (2004).

When two colinearly aligned inducing stimuli are present, one on each side, a boundary grouping can form even without direct bottom-up input: Long-range excitatory inputs converge onto the cell from both sides and summate. These excitatory inputs also activate a shared pool of inhibitory interneurons, which as well as inhibiting the target cell, also inhibit each other, thus normalizing the total amount of inhibition emanating from the interneuron pool. This summating excitation and normalizing inhibition create a case of "two-against-one" and the target cell is excited above-threshold (cf., von der Heydt and Peterhans, 1989; von der Heydt, Peterhans and Baumgartner, 1984). Finally, when there is direct bottom-up input, it can activate the cell without horizontal interactions.

Various modeling studies have suggested that the inhibitory interneurons may do more than realize the bipole grouping property. In particular, it has been proposed that inhibitory interneurons also inhibit the pyramidal cells that correspond to other orientations, notably perpendicular orientations. Here we propose that some of these inhibitory interneurons may also realize the disparity filter as part of the grouping process. Grossberg and Howe (2003) had suggested that the disparity filter needed to occur somewhere between layers 4 and 2/3 in V2. Our proposal puts it within layer 2/3. In other words, we propose that *the selection of a correct 3D grouping includes the suppression of false binocular matches.* Thus the hypothesis links a solution of the correspondence problem to the Gestalt grouping problem.

This hypothesis brings together several lines of psychophysical data. First, this hypothesis enables the precept model to simulate all of the stereopsis data that were explained in Grossberg and Howe (2003), as well as some data that were not explicable there. Second, this hypothesis explicates another important line of psychophysical data: It is often the case that the perceived depths of perceptual groupings covary with the disparities of the image contrasts from which they are generated. However, when this is not the case, then the perceived depth of emergent perceptual groupings can override local image disparities (Ramachandran and Nelson, 1976; Tausch, 1953; Wilde, 1950). Said in another way, a winning 3D grouping can suppress "false matches" that are based on the real local disparities of their generating features in the outside world. Placing the disparity filter within the inhibitory interneurons of the 3D grouping process can explain these data as well as how the 3D grouping process can suppress false matches that do not correspond to the correct image matches from the outside world.

(6) *Forms 3D surface percepts*. The above constraints all concern how the brain constructs a 3D boundary representation of an object. Much evidence suggests that boundary representations on their own do not give rise to visible percepts, which rather are a property of surface representations (Grossberg, 1994). Indeed, boundary representations pool opposite-polarity contrasts at the complex cell stage in order to be able to build boundaries of objects in front of textured backgrounds. Surface representations are proposed to derive from a filling-in process whereby lightness and color mark the depths at which the surfaces occur. Filling-in is

needed to recover lightness and color estimates in regions where they have been suppressed by the process of discounting the illuminant (Grossberg and Todorovic, 1988). Boundaries control the depths at which particular lightnesses and colors can fill-in, a process called 3D *surface capture*. The present model considers only the filling-in of achromatic lightness.

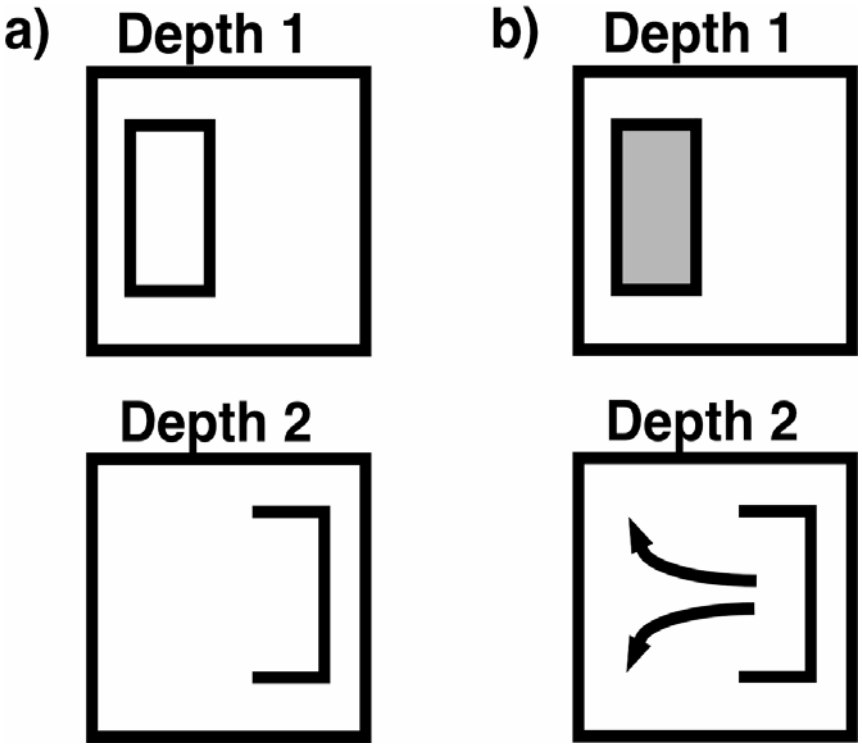


Figure 4. (a) Open and connected boundaries; (b) Filling-in of surface lightness is contained or not depending on the connectedness of the boundary. Note that the monocular boundaries (i.e., two horizontal boundaries and the right vertical boundary) have been added to all depth planes whereas the binocular boundary (i.e., the left vertical boundary) is present only in the near depth plane, thereby creating a connected boundary, and thus containment of filling-in, only in the near depth plane. [Reproduced with permission from Grossberg and Howe (2003).]

How does the brain ensure that lightness fills-in at only the correct depths? Grossberg (1994) proposed properties of this boundary-surface interaction that helped to explain many data about 3D figure-ground perception, as part of his development of FACADE theory, of which the 3D LAMINART model forms a part. In the 3D LAMINART model, one of these properties proved essential to explain 3D surface percepts that arise in stereopsis research. Namely, visible surfaces arise in cortical area V4 only if they are enclosed by *connected* boundaries. In particular, as diagrammed in Figure 4, a rectangular connected boundary may be composed of one vertical binocular boundary, one vertical monocularly viewed boundary, and two horizontal boundaries that code no disparity information. This connected boundary can support a visible surface percept at the depth corresponding to the binocular boundary if all other constraints are satisfied. Such a boundary can contain the filling-in process. However, if the vertical binocular boundary is missing, as it would be at a different depth plane, then the total boundary is not connected, and a visible percept will not be evident at that depth because filling-in can dissipate

out of the boundary gap. This example illustrates how the Monocular-Binocular Interface Problem (constraint (4) above), and thus the correspondence problem (constraint (3) above), influence visible percepts of 3D surfaces.

(7) Surface-to-boundary feedback ensures perceptual consistency and initiates figure-ground separation. Another property of FACADE theory is also needed to explain the stereopsis data that are the target of the present article; namely, that successfully filled-in regions within the surface representations send feedback to the boundary system. The surface system does this by sensing whether or not a surface region is filled-in by using contrast-sensitive output circuits that can detect where the bounding contours of the filled-in region occur. Such circuits can sense whether a surface region contains its internal lightness or color within a connected boundary, or allows it to dissipate because it does not possess a connected boundary. These contrast-sensitive output signals are realized by on-center off-surround networks that operate within disparity and across position within the surface system. The outputs from the surface stream to the boundary stream strengthen, and thereby confirm, the boundaries that surround the successfully filled-in surface regions, at the same time that they inhibit, or prune, redundant boundaries at the same positions and further depths (Grossberg, 1994, 1997; Grossberg and McLoughlin, 1997). This strengthening-and-pruning operation is accomplished by on-center off-surround networks that operate across disparity and within position within the boundary system. Surface-to-boundary feedback hereby assures that the boundary and surface representations are consistent. In the present model, the boundary-enhancing property interacts with a developmental bias that favors the fixation plane to explain stereopsis data that cannot otherwise be explained by the 3D LAMINART model.

The model laminar circuits that embody constraints 1-4 and 6 were used in Grossberg and Howe (2003) to simulate many challenging data about 3D vision, including: contrast variations of Panum's limiting case, dichoptic masking and the correspondence problem (McKee et al., 1994; Smallman and McKee, 1995), the Venetian blind illusion (Howard and Rogers, 1995), four different examples of da Vinci stereopsis (Gillam et al., 1999; Nakayama and Shimojo, 1990), stereopsis with opposite-contrast stimuli (Howard and Rogers, 1995; Howe and Watanabe, 2003; Julesz, 1971), the effect of interocular contrast differences on stereoacuities (Schor and Heckmann, 1989), and various lightness illusions. In so doing, it demonstrated the various roles that boundary and surface representations play in depth perception, and made a number of testable neurophysiological predictions. Here we show that constraints 5 and 7 are also needed in order to explain various additional data about 3D perception of closure and da Vinci stereopsis. Grossberg and McLoughlin (1997) first used surface-to-boundary feedback to explain one variant of da Vinci stereopsis, before the 3D LAMINART model was introduced. Grossberg and Howe (2003) did not need surface-to-boundary feedback to simulate several variants of da Vinci stereopsis, but did require a disparity filter that included cyclopean inhibition, as noted in Section (3). The present model shows how the reintroduction of surface-to-boundary feedback enables these and other variants of da Vinci stereopsis to be simulated, including variants that could not be simulated by Grossberg and Howe (2003), and does so without the use of cyclopean inhibition. An important part of this explanation concerns how constraints 5 and 7 expand the explanatory range of the 3D LAMINART model without undermining the explanations of the phenomena that do not seem to depend upon them. In other words, why does the dog bark only at the right times?

2. Model description

The enhanced 3D LAMINART model consists of five component networks which process: V1 monocular boundaries, V1 binocular boundaries, V2 boundaries, V2 monocular surfaces and V4 surfaces. Figure 5a shows a laminar cortical circuit diagram of the enhanced model. See Figure 5b for a block diagram of the model. A mathematical description of the model is elaborated in the Appendix. A preliminary version of the model was reported in Cao and Grossberg (2004) and Grossberg and Cao (2004). In order to reduce the computational load, the model currently considers only horizontal and vertical contours and five depth planes. Although model cells and cells in vivo will be clearly distinguished in the text, model cells will be referred to by physiological labels because their properties so closely match those found in vivo.

2.1. V1 monocular boundaries. The network that processes the V1 monocular boundaries comprises the monocular cells in layers 4, 3B and 2/3 of the V1 interblob region. The left and right retinal images are first processed by LGN cells which have circularly symmetric on-center off-surround receptive fields. These LGN cells discount the illuminant and enhance the scenic contrast. The monocular simple cells in layer 4 receive inputs from LGN cells. These simple cells act as oriented filters (Hubel and Wiesel, 1968), which are sensitive to either dark-light or light-dark contrast polarity, but not both. Layer 2/3 consists of complex cells, which add inputs from simple cells at the same position that are sensitive to the same orientation but opposite contrast polarities. These complex cells therefore respond to both contrast polarities and so can respond all along an object's boundaries even if its contrast polarity, with respect to the background, reverses as its boundary is transversed. Layer 2/3 implements contrast-invariant boundary detection. This layer also implements an early stage of perceptual grouping: Complex cells in layer 2/3 with collinear, coaxial receptive fields excite each other via long-range horizontal axons, which also give rise to short-range, disynaptic inhibition via inhibitory interneurons. This balance of excitation and inhibition helps to control grouping by implementing the bipole property (Grossberg, 1999a, 1999b; Grossberg, Mingilla and Ross, 1997; Grossberg and Raizada, 2000; Grossberg and Williamson, 2001). The boundary grouping process, together with contrast-invariant boundary detection, allows consistent and connected object boundaries to be formed even in response to noisy textured backgrounds.

2.2. V1 binocular boundaries. The network that processes V1 binocular boundaries is located in the interblob area and includes the binocular cells in layers 3B and 2/3. Layer 3B begins the process of stereoscopic fusion. As described in Section 2.1, inputs from the LGN activate monocular simple cells in layer 4 of the V1 interblob regions. Left and right eye monocular simple cells then conjointly activate binocular simple cells in layer 3B whose depth sensitivity is determined by the relative retinal disparity of the layer 4 monocular cells that project to them. These binocular simple cells in layer 3B are sensitive to just one contrast polarity because only layer 4 simple cells with the same contrast polarity project to a single layer 3B simple cell. This enables the *same-sign hypothesis* to be implemented, because layer 3B simple cells are selective for both binocular disparity and contrast polarity.

Layer 3B also contains inhibitory interneurons. The activity of a binocular simple cell is suppressed by these inhibitory interneurons if its left and right eye inputs differ too much in magnitude. These inhibitory interneurons thus ensure that the binocular simple cells act like the "obligate cells" of Poggio (1991), responding only when their left and right eye inputs are approximately equal in magnitude. These obligate cells help to solve the correspondence problem by ensuring that only similar stimuli with similar contrast amplitudes in the left and right eye retinal images are stereoscopically fused.

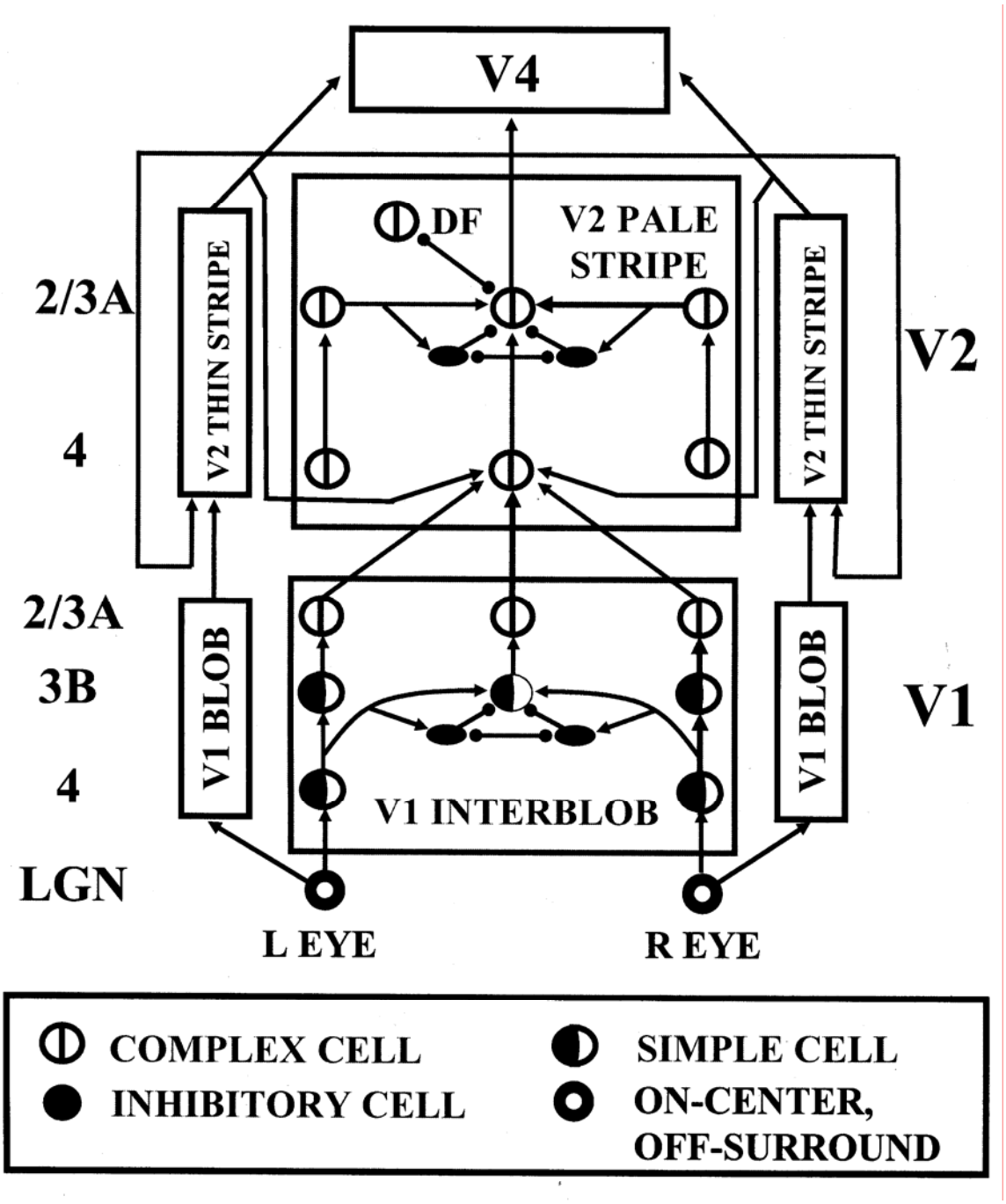


Figure 5 (a). The enhanced 3D LAMINART model circuit diagram. The model consists of a V1 Interblob - V2 Pale Stripe stream (Boundary Stream) which computes 3D perceptual groupings and a V1 Blob - V2 Thin Stripe stream (Surface Stream) which computes 3D surface representations of depth, lightness, and color. The two processing streams interact to overcome their complementary deficiencies and create consistent 3D boundary and surface percepts.

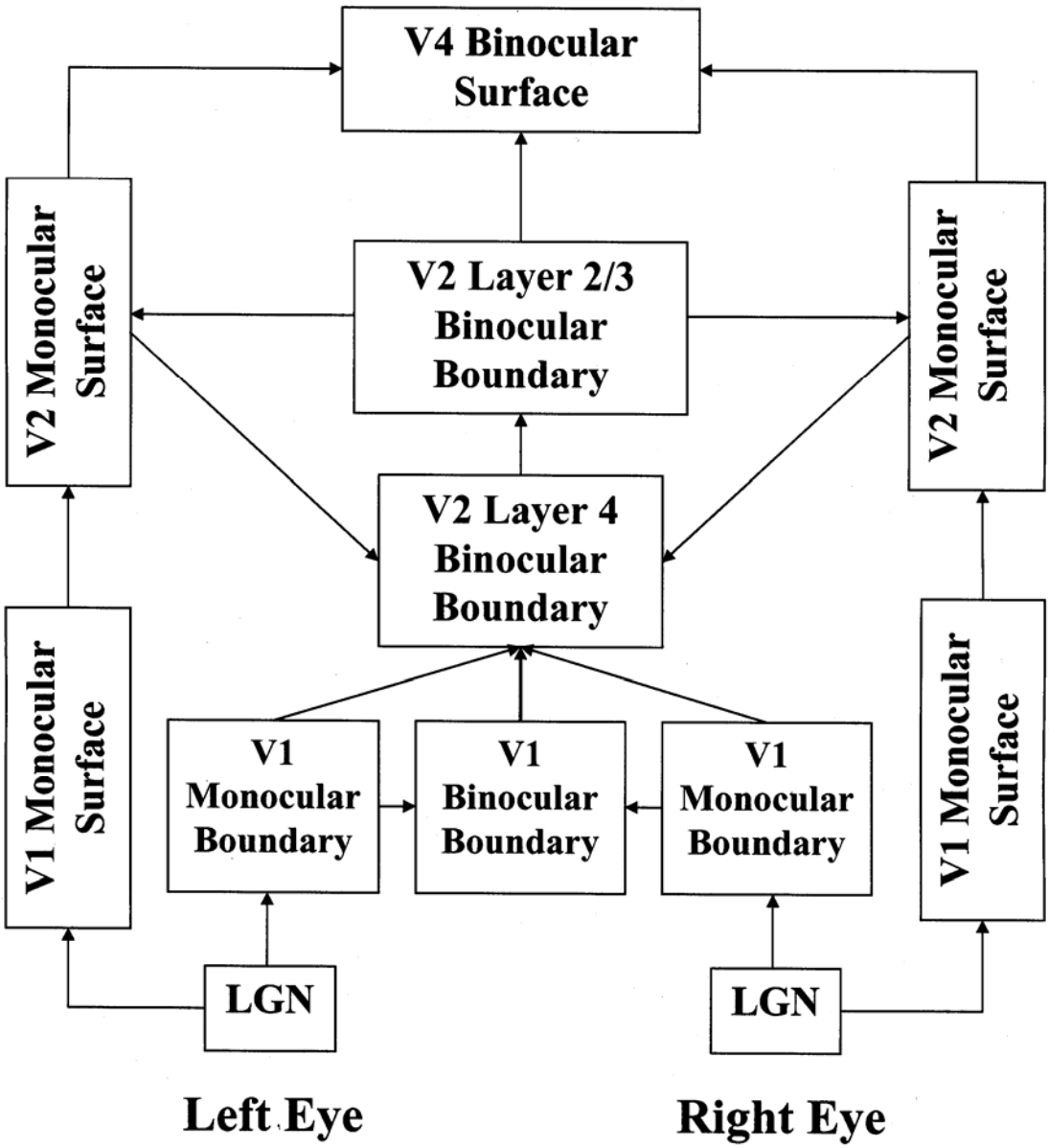


Figure 5 (b). A block diagram of the enhanced 3D LAMINART model.

Layer 3B simple cells that are sensitive to the same position and disparity but *opposite* contrast polarities pool their signals at layer 2/3 binocular complex cells. As in the monocular cases in Section 2.1, layer 2/3 complex cells implement contrast-invariant boundary detection and perceptual grouping, but these groupings are disparity-selective.

2.3. V2 boundaries. The network that processes the V2 boundaries is located in the V2 pale stripe region and includes the binocular complex cells in V2 layers 4 and 2/3. Monocular and binocular inputs are combined in V2 layer 4. In other words, V2 layer 4 cells receive inputs from left and right monocular complex cells and binocular complex cells in V1 layer 2/3. Since the monocular cells are not associated with a particular depth plane, their outputs are added to all depth planes in layer 4 along their respective lines-of-sight (see the Appendix). The layer 4 cells also receive feedback signals from left and right V2 monocular surfaces that are formed in the

V2 thin stripe region. These surface-to-boundary feedback inputs modulate corresponding V2 layer 4 cells in the following way: the activity of an active layer 4 cell is enhanced if it receives either a left or right surface-to-boundary excitatory feedback signal, or both. Its activity is suppressed otherwise. As shown in our simulations in Section 3, these surface-to-boundary feedback signals play an indispensable role in explaining percepts of some stereo displays.

V1 layer 3B binocular cells attempt to match every vertical edge in one retinal image with every other nearby vertical edge in the other retinal image within its disparity range that has the same contrast polarity and approximately the same magnitude of contrast. As a result, false matches may exist in V1. Figure 3 shows four possible matches if each eye sees two bars. Only the two matches in the fixation plane are correct, and the other two are false. Such false matches are suppressed in V2 via a disparity filter. Figure 3 illustrates how the disparity filter works. To encourage unique-matching, the model assumes that each neuron inhibits all other neurons that share either of its monocular inputs; that is, shares one of its monocular lines-of-sight. This is represented by the solid lines between neurons in Figure 3. Furthermore, the disparity filter is symmetrical about the fixation plane (i.e., the near and far disparity planes equally inhibit and are equally inhibited by the fixation plane). The disparity filter favors the fixation plane in that this plane inhibits the near and far disparity planes more than they inhibit it (see Table 2). The line-of-sight inhibition and the fixation plane advantage work together to ensure that the two matches in the fixation plane typically win, thereby helping to solve the correspondence problem. The model assumes that the disparity filter can be realized as part of the inhibitory interactions that control perceptual grouping by long-range horizontal connections in V2 layer 2/3, and thereby combines suppression of false matches with long-range Gestalt grouping processes (see Figure 5a and equations (A30)-(A35) in the Appendix).

2.4. V2 monocular surfaces. The network that processes the V2 monocular surfaces is located in the V2 thin stripe region. The left (right) V2 thin stripe receives boundary signals from V2 layer 2/3 complex cells and lightness signals from left (right) LGN cells via the left (right) V1 blob region (see Figure 5a). It has been proposed how surface representations may be generated by a filling-in process (Grossberg, 1994; Grossberg and Todorovic, 1988). Psychophysical data (e.g., Paradiso and Nakayama, 1991; Pessoa and Neumann, 1998; Pessoa, Thompson and Noe, 1998) and neurophysiological data (e.g., Lamme, Rodriguez-Rodriguez and Spekreijse, 1999; Rossi, Rittenhouse and Paradiso, 1996) support the existence of such a filling-in process. Surface filling-in has been used to explain many psychological phenomena, such as figure-ground separation (Kelly and Grossberg, 2000), 2D and 3D neon color spreading (Grossberg, 1994; Grossberg and Mingolla, 1985a; Grossberg and Swaminathan, 2004), and lightness perception (Grossberg and Hong, 2004; Grossberg and Kelly, 1999; Grossberg and Todorovic, 1988). As illustrated in Section 1 (item 5), a surface presentation can rise from filling-in only if it is enclosed by a connected boundary.

The LGN cells, which obey cell membrane equations and interact through on-center, off-surround circularly symmetric receptive fields, first discount the effects of a spatially non-uniform illumination; the excitatory and inhibitory components of these receptive fields are balanced so that cell responses are attenuated to spatially uniform or slowly varying stimulation. Model LGN cells therefore respond preferentially to luminance borders. At the later filling-in stage, these luminant-discounted border signals propagate throughout surface regions that are completely enclosed by boundaries to complete the lightness representation. The filling-in process has often been modeled by a boundary-gated diffusion equation (Grossberg and Todorovic, 1988), as defined in the Appendix. Propagating signals can dissipate across space

unless the region is surrounded by a connected boundary which creates resistive barriers to limit signal spread (Figure 4). The connected boundaries received from layer 2/3 act as such barriers and, together with the lightness information received from LGN, can create monocular surface representations in V2 thin stripes via the filling-in process. Successfully filled-in monocular surfaces are predicted to send contour-sensitive surface-to-boundary feedback signals into V2 layer 4 (Figure 5b). These surface-to-boundary signals modulate the activities of V2 boundary cells so that the boundaries that surround the successfully filled-in surfaces are enhanced and other boundaries are suppressed. See the Appendix for mathematical details.

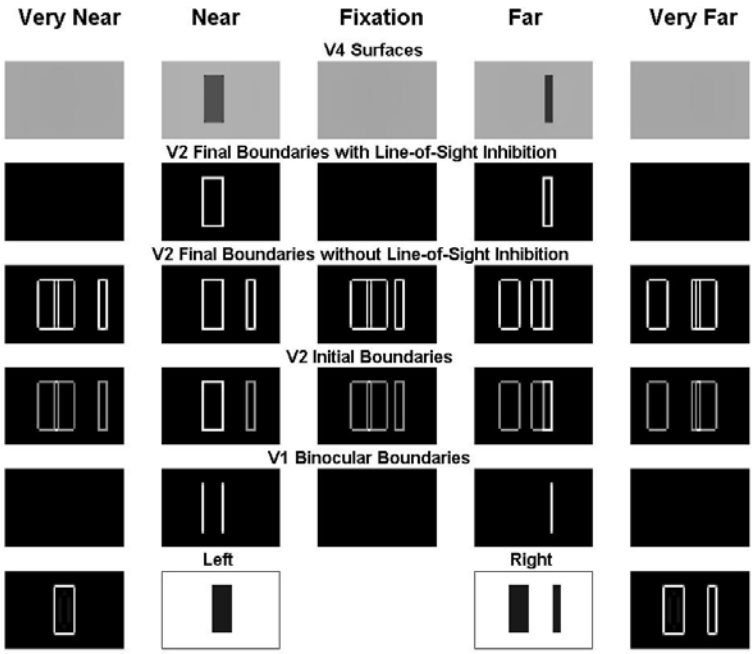
2.5. V4 surfaces. Although monocular surfaces are formed in V2 thin stripes, they are predicted to be invisible there and hereby do not create visible 3D surface percepts. The visible 3D surface percepts are proposed to be generated in cortical area V4. Area V4 receives boundary signals from V2 layer 2/3 and lightness signals from the LGN via V1 blobs and V2 thin stripes. The surface filling-in process is similar to the one described in Section 2.4, except V4 combines monocular lightness signals from both eyes.

3. Model simulations

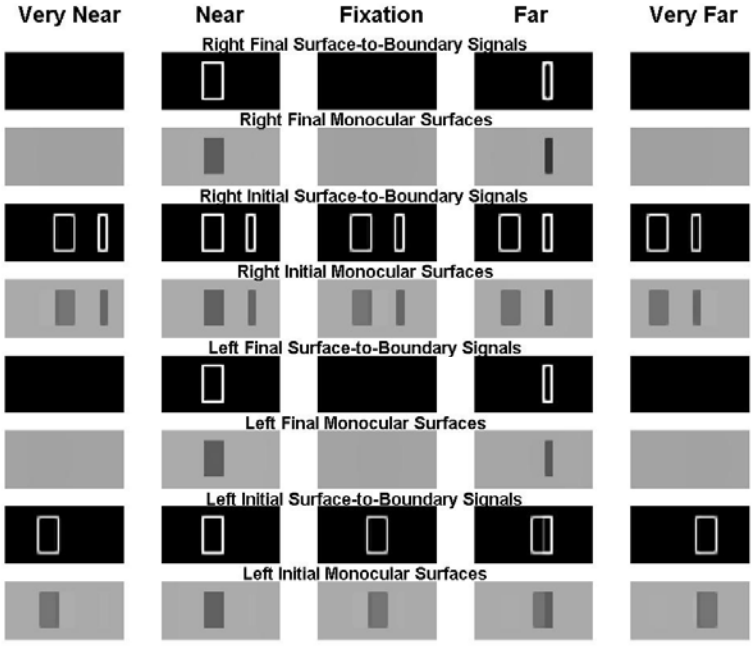
This section summarizes simulations that predict how monocular and binocular information interact in the visual cortex to generate 3D surface percepts. As described in Section 1, one of main new developments of the current model is that it clarifies how surface-to-boundary feedback plays an essential role in this process. Below two such cases are considered: Variations of da Vinci stereopsis and an example of perceptual closure. Simulations also show the surface-to-boundary feedback does not undermine explanations of any of the data that were simulated in Grossberg and Howe (2003). This cannot be taken for granted because surface-to-boundary signals can change the relative strength of boundaries across depth and thus the depthful surfaces that are perceived.

Each simulation figure includes two parts: (I) V1 and V2 boundaries and V4 surfaces and (II) V2 monocular surfaces and surface-to-boundary feedback signals. Like the model diagram shown in Figure 5a, Part (I) of each figure should be read from the bottom up, with the bottom two rows representing the input and the V1 boundary representations, the next two or three rows representing the V2 boundary representations, and the top row representing the V4 surface representations. In Part (II) of each figure, the bottom four rows represent, respectively, the initial and final V2 monocular surfaces and feedback signals for the left eye. The top four rows represent the same quantities for the right eye. Furthermore, for the top four rows of Part (I) and all rows of Part (II), depth increases from left to right, with the middle plot representing the fixation plane, the two leftmost plots representing the two near depth planes, and the two right plots representing the two far depth planes.

3.1. Variations of da Vinci Stereopsis. In the set of experiments of Nakayama and Shimojo (1990), a thick bar was presented to both eyes and a thin bar only to the right eye, as shown in the middle two plots of the first row of Figures 6a and 7a. Subjects reported perceiving the thin bar behind the thick bar, at a depth that was consistent with the right edge of the thin bar of the right input being fused with the right edge of thick bar of the left input. Grossberg and Howe (2003) explained the experiments where the thin bar fusion occurred at a far depth plane (see the fourth plot in the second row in Figure 6a). Here we consider a variation where the right edge of the thin bar of the right input was fused with the right edge of thick bar of the left input at the fixation plane, while the left edge of the thin bar of the right input was fused with the left edge of thick bar of the left input at a far depth plane (see the second row in Figure 7a). This case



(a)



(b)

Figure 6. Simulation of the original da Vinci stereopsis display explained by Grossberg and Howe (2003) and reported experimentally by Nakayama and Shimojo (1990). (a) The first row represents the inputs (the middle two plots) and the V1 monocular boundaries (the outer two

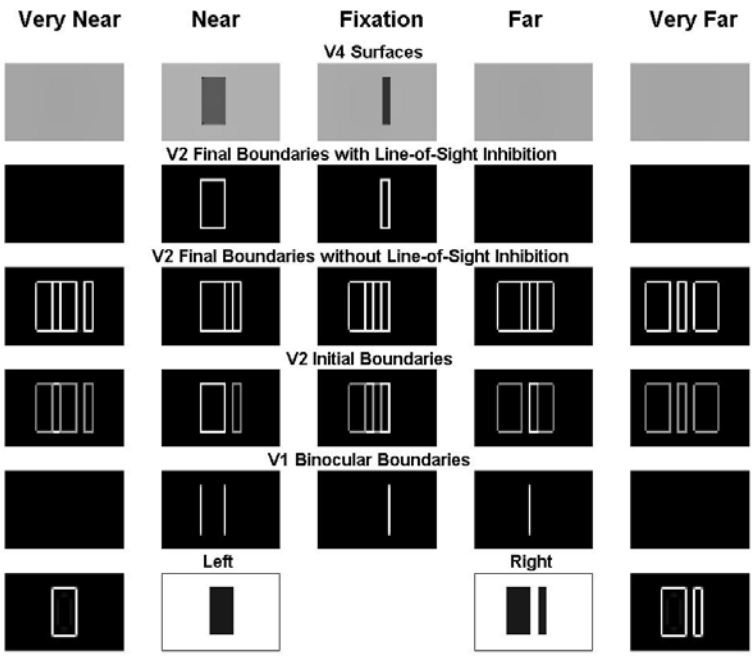
plots), the second row the V1 binocular boundaries, the third row the V2 layers 4 and 2/3 initial boundaries, the fourth row the V2 layer 2/3 final boundaries when line-of-sight inhibition in the disparity filter is shut off (that is, all coefficients in Table 2 are set equal to zero), the fifth row the V2 layer 2/3 final boundaries in the full model, and the sixth row the V4 surface percepts. In the top four rows, depth increases from left to right for successive plots, with the middle plot representing the fixation plane. (b) The bottom four rows represent, respectively, the initial and final V2 monocular surfaces and surface-boundary feedback signals for the left eye. The top four rows represent the same quantities for the right eye. In all rows, depth increases from left to right for successive plots, with the middle plot representing the fixation plane. Increasing boundary strength is denoted by whiter lines at the boundary positions, since boundaries are plotted against a black background. More intense filling-in, and thus greater contrast, of the filled-in surface regions is denoted by darker or whiter surface regions on a grey background. All simulation plots use a similar format.

will happen when moving the thin bar of the right input closer to the thick bar than in the original case shown in Figure 6a.

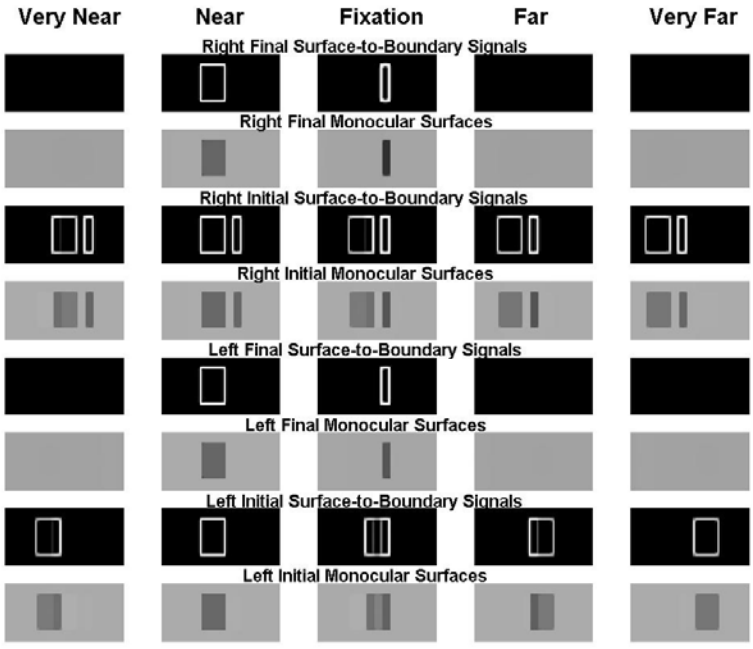
Let us first see what the model in Grossberg and Howe (2003) would predict for this new case. Figure 8 shows the simulation result by the Grossberg and Howe (2003) model. The vertical boundaries of the thick bar are registered binocularly in the near disparity plane in V1, as shown by the second plot of the second row of Figure 8. The right edge of the thin bar is matched with the right edge of the thick bar to be registered binocularly in the fixation plane in V1, as shown by the third plot. Finally, the left edge of the thin bar is matched with the left edge of the thick bar to be registered binocularly in the far disparity plane in V1, as shown by the fourth plot.

The monocular boundaries (displayed in the leftmost and rightmost plots of the first row) are added to all depth planes in the V2 disparity filter along their respective monocular lines-of-sight, as shown by the plots in the third row of Figure 8. The binocular vertical boundaries in V1 (displayed in the second row) are also added to the disparity filter, overlapping with the vertical boundaries of the thick bar representation in the second plot, with the rightmost vertical boundary in the third plot, and with the third vertical boundary in the fourth plot. These vertical boundaries, being much stronger, eliminate all other weaker vertical boundaries that share their lines-of-sight via the disparity filter line-of-sight inhibition. As a result, the left edge in the fixation plane, the right edge in the far depth, and all vertical edges in other depths of the thin bar representations are eliminated, as shown in the fourth row. The thin bar thus has no connected boundary representation in any depth plane.

Since surfaces can successfully be filled in only for those regions that are completely enclosed by a connected boundary, the surface representation of the thin bar disappears in all depth planes. The thick bar boundaries in the near depth are much stronger and hereby not eliminated. The connected thick bar boundaries in the near depth can form a surface by filling-in. The final percept is shown in the top row. The thick bar appears in the near depth and the thin bar disappears. Note that the left vertical boundary in the third plot (the fixation plane) of the fourth row is not from the left edge of the thin bar. In contrast, it is from the right edge of the thick bar of the right input. Therefore, the surface created by the connected boundaries in the third plot, as shown in the third plot of the top row, is not the true surface representation of the thin bar. In fact, it is obviously much wider than the thin bar. The Grossberg and Howe (2003)



(a)



(b)

Figure 7. Simulation of a variation of the da Vinci stereopsis display of Figure 6. Surface-to-boundary feedback plays a critical role in determining the final percept. See text for details.

model thus fails to make the correct prediction. The reason is that the left vertical boundary representation of the thin bar in the fixation plane is eliminated by its corresponding stronger boundary representation in the far depth plane.

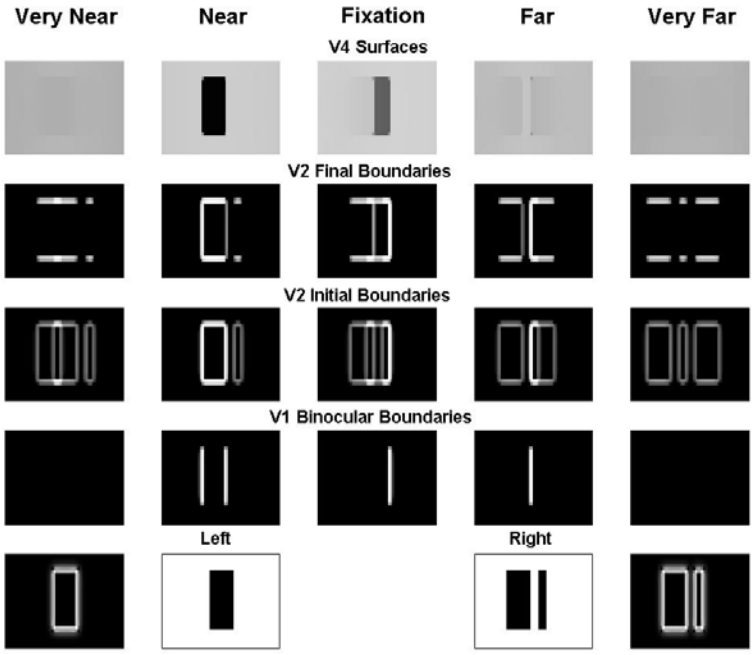


Figure 8. Simulation of the da Vinci stereopsis display of Figure 7 by the Grossberg and Howe (2003) model.

Now let us see what the enhanced 3D LAMINART model that contains surface-to-boundary feedback would predict. The simulation is shown in Figure 7. The initial stages (the bottom three rows of Figure 7a) of the model are basically the same as of the Grossberg and Howe (2003) model. The boundaries of the thick bar are registered binocularly in the near disparity plane in V1, the right edge of the thin bar is matched with the right edge of the thick bar to be registered binocularly in the fixation plane, and the left edge of the thin bar is matched with the left edge of the thick bar to be registered binocularly in the far disparity plane, as shown by the second, third and fourth plots of the second row of Figure 7a, respectively. The monocular boundaries (displayed in the leftmost and rightmost plots of the first row) are added to all depth planes in V2 layer 4 along their respective monocular lines-of-sight, as shown by the plots in the third row of Figure 7a. The binocular boundaries in V1 (displayed in the second row) are also added to the V2 layer 4 in their own disparity planes. These boundaries in V2 layer 4 then input into V2 layer 2/3. The V2 layer 2/3 acts as the disparity filter to solve the correspondence problem while also controlling perceptual grouping by horizontal connections. Furthermore, the boundaries in V2 layer 2/3 control the formation of left and right eye monocular surfaces in the V2 thin stripes, together with the inputs from the left and right LGNs. These monocular surfaces that are initially formed are shown in rows one and five of Figure 7b. These surfaces fill-in a subset of the initial boundaries shown in row three of Figure 7a, namely all the closed connected boundaries that receive contrastive surface inputs.

The monocular surfaces in the V2 thin stripes then send surface-to-boundary feedback signals, as shown in rows two and six of Figure 7b, to V2 layer 4 to modulate the activities of the corresponding boundary cells. The cells that receive surface-to-boundary feedback signals from left, right, or both eyes are enhanced, while cells that do not receive such surface-to-boundary signals are suppressed. Initially, the V2 boundaries that receive binocular inputs from V1 − namely, the vertical boundaries of the thick bar representation in the second plot of the third row, the rightmost vertical boundary in the third plot, and the third vertical boundary in the fourth plot − are much stronger than the others. These stronger boundaries would ordinarily eliminate other weaker boundaries that share their lines-of-sight via the disparity filter line-of-sight inhibition. However, they do not eliminate the vertical boundary originating from the left edge of the thin bar in the fixation plane due to the boundary-enhancing effect of surface-to-boundary feedback signals.

The reasons are as follows: When the right vertical boundary representation (weaker one) of the thin bar in the far depth plane is eliminated, the left vertical boundary representation (weaker one) of the thin bar in the fixation plane still remains, since the fixation plane is favored. In other words, the disparity filter favors the fixation plane because this plane inhibits the near and far disparity planes more than they inhibit it. This property has proved to be essential in helping to explain many stereopsis phenomena, such as contrast variations of the correspondence problem that were studied by Smallman and McKee (1995) (see Grossberg and Howe, 2003). Once the right vertical boundary representation of the thin bar in the far plane is eliminated, its connected boundary representation in the far plane is destroyed. As a result, the V2 monocular surface of the thin bar that is associated with the far disparity plane cannot be formed, since the surface filling-in will then flow out (see the fourth plot in the seventh row of Figure 7b). As a result, the corresponding boundary cells in V2 layer 4 do not receive surface-to-boundary signals and hence become weaker compared to those boundary cells that do receive surface-to-boundary signals. This effect makes the left vertical boundary representations of the thin bar become stronger in the fixation plane and weaker in the far disparity plane. As a result, the left vertical boundary representation of the thin bar in the fixation plane is not eliminated. The final V2 boundary representations are shown in the fifth row of Figure 7a. In comparison, the fourth row shows the result when line-of-sight inhibition in the disparity filter is shut off. As explained before, V4 fills in surfaces in those regions that are completely enclosed by a connected boundary. This produces a percept of a thick bar in a near disparity plane, represented by the second plot of the top row, and a thin bar in the fixation plane, represented by the third plot. The model therefore correctly predicts that the thin bar will appear behind the thick bar at a depth that is consistent with the right edge of the thin bar being stereoscopically fused with the right edge of the thick bar, as has been reported experimentally (Nakayama and Shimojo, 1990). The model hereby proposes how the surface-to-boundary feedback can play an essential role in making the correct prediction.

The enhanced 3D LAMINART model also consistently explains the original da Vinci Stereopsis case reported in Grossberg and Howe (2003). It is essential to realize how this happens, since it is as important to know when a mechanism does not change a percept as when it does. Why this "dog does not bark" is explained in the simulation summarized in Figure 6. The vertical boundaries of the thick bar are registered binocularly in the near disparity plane in V1, as shown by the second plot of the second row, and the right edge of the thin bar is matched with the right edge of the thick bar to be registered binocularly in the far disparity plane in V1, as shown by the fourth plot. The left edge of the thin bar is registered only monocularly because it

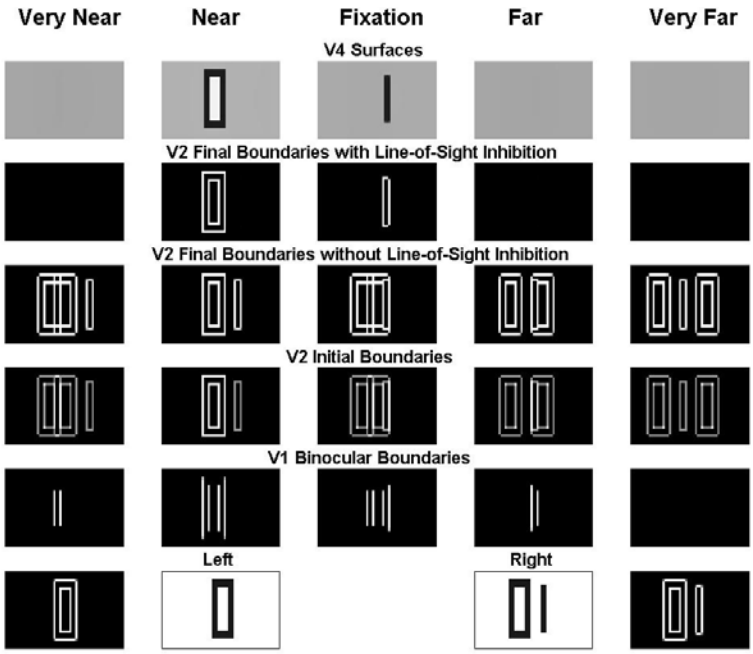
cannot be matched with either of the edges of the left input. As usual, the monocular boundaries are added to all depth planes in V2 layer 4 along their respective monocular lines-of-sight, as shown by the plots in the third row of this figure. The vertical binocular boundaries are also added to V2 layer 4, overlapping with the vertical boundaries of the thick bar representation in the second plot and with the rightmost vertical boundary in the fourth plot.

All boundaries in V2 layer 4 then input into V2 layer 2/3. Initially, the V2 layer 2/3 vertical boundaries which receive binocular inputs from V1 − namely, the vertical boundaries of the thick bar representation in the second plot of the third row and the rightmost vertical boundary in the fourth plot − are stronger and thus eliminate all other weaker vertical boundaries that share their lines-of-sight via the disparity filter line-of-sight inhibition. As a result, the boundary representation of the right edge of the thin bar in any depth plane other than the far depth is eliminated. Once this has happened, then only the far depth plane has connected boundaries of the thin bar and thus can continue to form a monocular surface via filling-in (see the seventh row of Figure 6b). In other words, the monocular surface representations of the thin bar that are associated with other depth planes cannot be filled in and hereby cannot send surface-to-boundary feedback signals anymore. Therefore, the boundary representations of the left edge of the thin bar in all other depths than the far depth are depressed because they do not receive surface-to-boundary feedback signals. In particular, the left vertical boundary of the thin bar in the fixation plane is depressed. Therefore, even though the fixation plane is favored, the left vertical boundary of the thin bar in the far depth plane can win the competition.

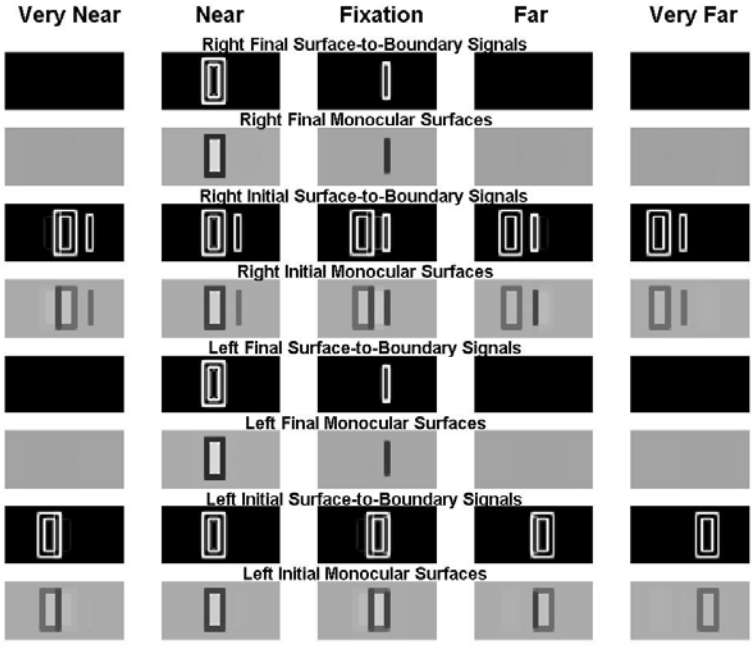
The final V2 boundary representations are shown in the fifth row of Figure 6a. In comparison, the fourth row shows the result when line-of-sight inhibition in the disparity filter is shut off. As usual, V4 fills in surfaces in those regions that are completely enclosed by a connected boundary. This produces a percept of a thick bar in a near disparity plane, represented by the second plot of the top row, and a thin bar in a far disparity plane, represented by the fourth plot. The model therefore correctly predicts that the thin bar will appear behind the thick bar at a depth that is consistent with the right edge of the thin bar being stereoscopically fused with the right edge of the thick bar, as has been reported experimentally by Nakayama and Shimojo (1990).

3.2. Perceptual Closure. The closure case cannot be explained by the Grossberg and Howe (2003) model. Here, a rectangular frame is presented to both eyes and a single bar with the same thickness as the edges of the rectangular frame is presented only to the right eye, as shown in the middle two plots of the first row of Figure 9a. The single bar lies beside the rectangular frame at a distance such that the left and right edges of the rectangular frame of the left input respectively fuse with the right edge of the rectangular frame of the right input and the single bar. Subjects reported perceiving the rectangular frame in the front of a single bar that is seen in the fixation plane.

Figure 10 shows by simulation why the Grossberg and Howe (2003) model cannot explain this result. The vertical boundaries of the rectangular frame are fused binocularly in a near disparity plane in V1, as shown by the second plot of the second row. The left edge of the rectangular frame of left input is matched with the right edge of the rectangular frame of right input and is registered binocularly in the fixation plane, as shown by the leftmost pair of vertical boundaries in the third plot. The right edge of the rectangular frame of the left input is fused binocularly with the single bar of the right input in the fixation plane, as shown by the rightmost pair of vertical boundaries in the third plot. The left edge of the rectangular frame of left input is matched binocularly with the single bar in a far depth plane, as shown by the fourth plot, and the



(a)



(b)

Figure 9. Simulation of perceptual closure case. Surface-to-boundary feedback plays a critical role in determining the final percept. See text for details.

right edge of the rectangular frame of the left input is matched binocularly with the left edge of the rectangular frame of the right input in a very near depth plane, as shown in the first plot. As usual, the monocular boundaries are added to all depth planes in the disparity filter in V2 along their respective monocular lines-of-sight, as shown by the plots in the third row of Figure 10. The vertical binocular boundaries in V1 (displayed in the second row) are also added to the disparity filter in their own disparity planes, overlapping with the vertical boundaries of the rectangular frame representation in the second plot of third row, with the rightmost two pairs of vertical boundaries in the third plot, with the third pair of vertical boundaries in the fourth plot, and with the second pair of vertical boundaries in the first plot.

These vertical boundaries, being much stronger, quickly eliminate all other weaker vertical boundaries that share their lines-of-sight via the disparity filter line-of-sight inhibition. At the same time, the stronger vertical boundaries in different depth planes can also inhibit each other. Since the fixation plane is favored, the rightmost two pair of vertical boundaries in the fixation plane eliminate all vertical boundary representations in other depth planes which share either line-of-sight with them, as shown in the fourth row. In particular, the vertical boundary representations of the rectangular frame in the near depth plane are eliminated, as shown in the second plot of the fourth row. As a result, surfaces cannot be filled in there. The final prediction is shown in the top row. The rectangular frame disappears. Therefore, the Grossberg and Howe (2003) model fails to make the correct prediction. It is shown now how the correct percept obtains when the surface-to-boundary feedback process is included.

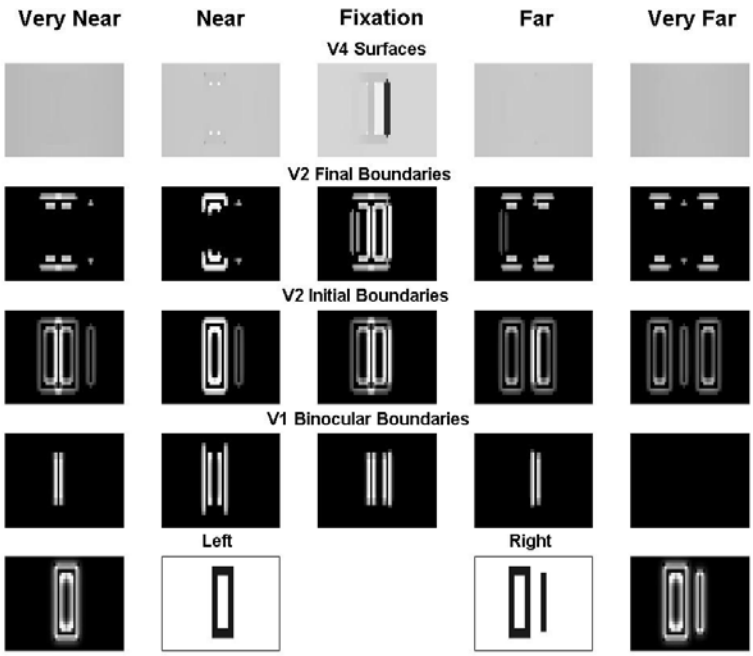


Figure 10. Simulation of perceptual closure case by Grossberg and Howe model.

Figure 9 shows the simulation by the enhanced model. The initial stages (the bottom three rows of Figure 9a) of the model are basically the same as those of the Grossberg and Howe (2003) model. The boundaries of the rectangular frame are fused binocularly in a near disparity plane in V1, as shown by the second plot of the second row, while the left and right edges of the

rectangular frame of the left input are, respectively, matched with the right edge of the rectangular frame of right input and the single bar to be registered binocularly in the fixation plane, as shown by the third plot. The binocular fusions appearing in the very near depth plane and the far depth plane, as shown by the first and fourth plots, are less essential. As usual, the monocular boundaries are added to all depth planes in the V2 layer 4 along their respective monocular lines-of-sight, as shown by the plots in the third row of this figure. The binocular vertical boundaries in V1 are also added to the V2 layer 4 in their own disparity planes. All boundaries in V2 layer 4 then input into V2 layer 2/3. As described before, V2 layer 2/3 acts as the disparity filter to solve the correspondence problem while also controlling perceptual grouping by horizontal connections, the boundaries in V2 layer 2/3 control the filling-in of left and right monocular surfaces in the V2 thin stripes, and the successfully filled-in monocular surfaces in the V2 thin stripes send surface-to-boundary feedback signals to V2 layer 4 to modulate the activities of the corresponding boundary cells.

This modulation makes the cells that receive positive surface-to-boundary feedback signals get enhanced and other cells that do not receive surface-to-boundary signals get suppressed. In particular, the stronger vertical boundaries in V2 layer 2/3 eliminate other weaker boundaries that share their lines-of-sight via disparity filter line-of-sight inhibition. Since the fixation plane is biased, the boundary representations of the rectangular frame in any depth plane other than the fixation plane would be eliminated if there were no surface-to-boundary feedback modulation. However, because of the effect of surface-to-boundary modulation, the boundary representations of the rectangular frame in the near depth plane are not eliminated. The reason is that the strong connected boundary cells of the rectangular frame in the near disparity plane cause strong monocular surfaces to fill-in within the V2 thin stripes (see the second plots in the odd rows of Figure 9b), which then send strong surface-to-boundary feedback signals (see the second plots in the even rows of Figure 9b) to V2 layer 4 to enhance the corresponding boundary cells there. As a result, the boundary representations of the rectangular frame in the near depth plane become stronger and hence are not eliminated. The final V2 boundary representations are shown in the fifth row of Figure 9a. In comparison, the fourth row shows the result when line-of-sight inhibition in the disparity filter is shut off. As usual, V4 fills in surfaces in those regions that are completely enclosed by a connected boundary. This produces the correct percept of a rectangular frame in a near disparity plane, represented by the second plot of the top row, and a single bar in the fixation plane, represented by the third plot, in Figure 9a.

3.3. Cases explained in Grossberg and Howe (2003). The enhanced 3D LAMINART model can also explain all cases discussed in Grossberg and Howe (2003). It offers explanations as good as those in Grossberg and Howe (2003) for stereopsis cases such as contrast variations of dichoptic masking and the correspondence problem, the effect of interocular contrast differences on stereoacuity, Panum's limiting case, and the Venetian blind illusion. In all these cases, stronger boundaries receive stronger surface-to-boundary feedback signals, since they can better generate surfaces through filling-in. This feedback makes the stronger boundaries even stronger. They hereby eliminate weaker boundaries more efficiently via disparity filter line-of-sight inhibition. Therefore, the final results are the same as those in Grossberg and Howe (2003).

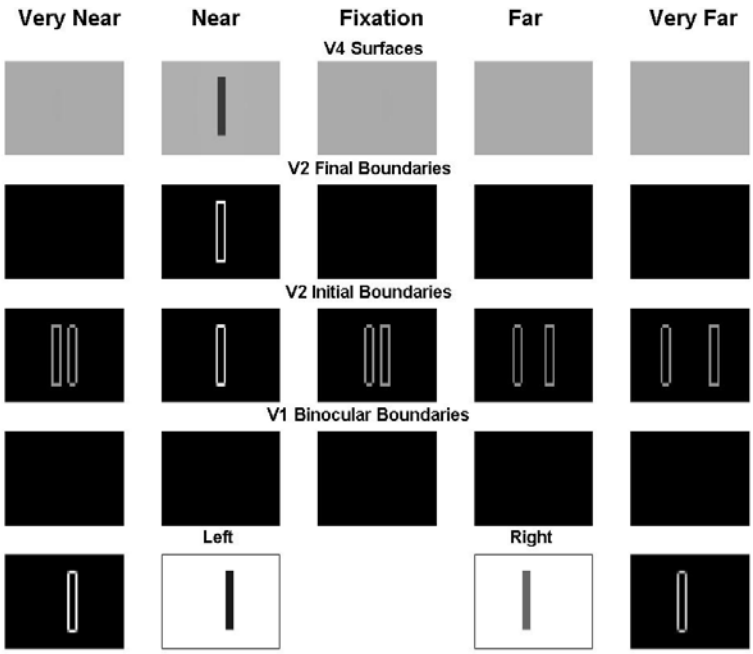
The enhanced model can also produce consistent explanations for the da Vinci stereopsis experiments in Gillam et al. (1999) and the stereopsis cases with polarity-reversed stereograms that were discussed in Grossberg and Howe (2003). In all these cases, there is at least one input bar and one depth plane in which one vertical boundary of the input bar is binocularly fused and the other vertical boundary of the input bar is only monocularly viewed, while in other depth

planes, both vertical boundaries of the input bar are monocularly viewed. Since binocularly viewed boundaries are stronger than monocularly viewed boundaries, disparity filter line-of-sight inhibition may cause the boundary representations for one vertical edge of the input bar to be eliminated in some depth planes while the boundary representations for the other vertical edge remain. These remaining redundant vertical boundaries may hinder the correct percept in some cases (e.g., the variations of da Vinci stereopsis discussed in Section 3.1). The surface-to-boundary feedback can suppress the remaining redundant boundaries and hereby help to create the correct prediction.

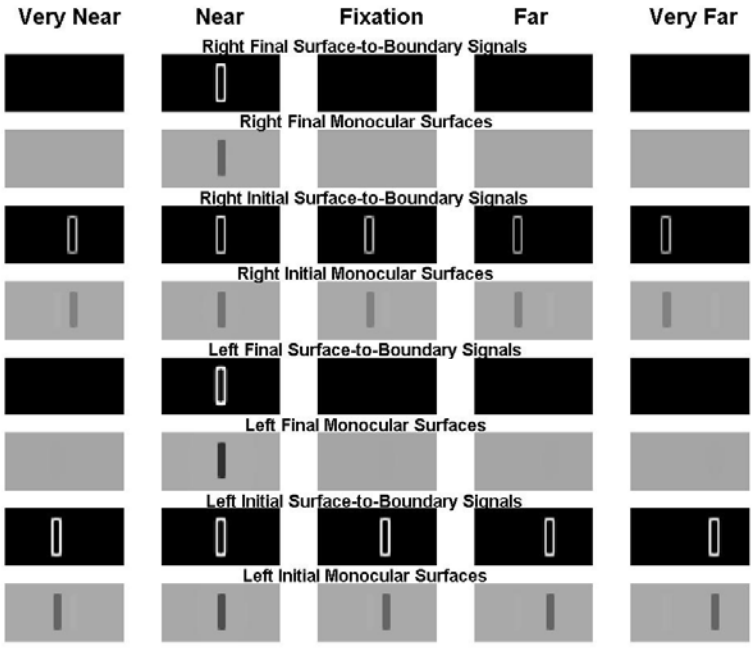
3.3.1. Dichoptic masking. In the basic paradigm considered by McKee et al. (1994), a high contrast bar is presented to the left eye and a low contrast bar is presented to the right eye, as shown in the middle two plots in the bottom row of Figure 11a. Subjects reported perceiving that the high contrast bar masks the low contrast bar. The model explanation is as follows: Since their contrasts differ too much, these two bars cannot binocularly fuse in V1. As usual, monocular boundaries are added to all depth planes in V2 layer 4 along their respective monocular lines-of-sight, as shown in the third plot of Figure 11a. Note that the left and right monocular boundaries coincide in the near depth plane represented by the second plot of this row, and vertical boundaries in this depth plane are hereby stronger than those in other depth planes. All boundaries in V2 layer 4 then input into V2 layer 2/3, which controls the monocular surface filling-in within the V2 thin stripes.

The successfully filled-in monocular surfaces then send surface-to-boundary signals to V2 layer 4. Initially all monocular surfaces are successfully filled-in. The boundaries in the near depth plane, being stronger, create stronger V2 monocular surfaces and thus stronger surface-to-boundary feedback signals. As a result, the vertical boundaries in the near depth plane remain stronger than others after the surface-to-boundary feedback. They then eliminate boundaries in other depth planes via disparity filter line-of-sight inhibition. The final V2 boundary representations are shown in the fourth row of Figure 11a. As usual, V4 fills in surfaces in those regions that are completely enclosed by a connected boundary. This produces a percept of a single bar in a near disparity plane, represented by the second plot of the top row, as has been reported experimentally (McKee et al., 1994).

Based on the same reasoning, the enhanced model can equally well explain other dichoptic masking cases including the Panum's limiting case discussed in Grossberg and Howe (2003). In all these cases, surface-to-boundary feedback does not undermine the relative strength of V2 boundaries. The simulations that demonstrate this claim are shown in Figures 12-15. Figure 12 shows the model simulation of the release from dichoptic masking studied by McKee et al. (1995). Here a high contrast bar is presented to both eyes and a low contrast bar is presented to only the right eye. The model correctly predicts that the low contrast bar is released from masking by the high contrast bar, as reported experimently by McKee et al. (1995). Figure 13 show the model simulation of a variation of the release from dichoptic masking studied by McKee et al. (1995). Here a low contrast bar is presented to both eyes and a high contrast bar is presented to only the left eye. The model again correctly predicts that the low contrast bar is released from masking by the high contrast bar. Figure 14 show the model simulation of the return to dichoptic masking studied by McKee et al. (1995). Here a high contrast bar is presented to the left eye and two low contrast bars are presented to the right eye. The model correctly predicts that the high contrast masks the left bar of the right input again. Figure 15 shows the model simulation of dichoptic masking in Panum's limiting case studied by McKee et al. (1995). Here all three bars have the same contrast. The model correctly predicts that the left input bar is

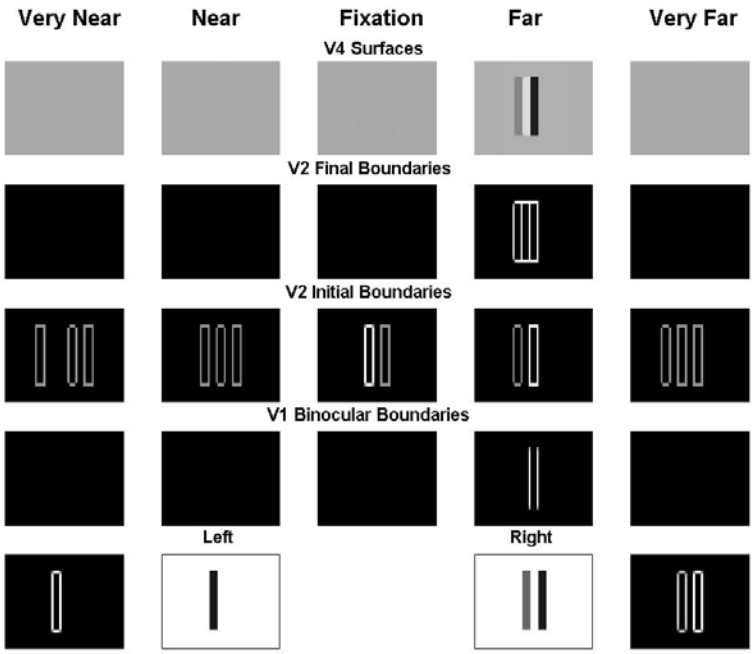


(a)

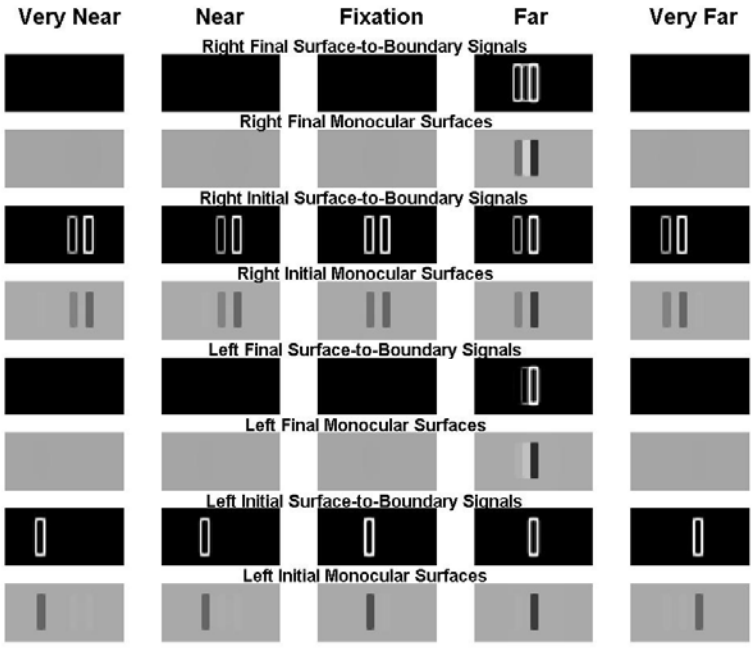


(b)

Figure 11. Simulation of the basic dichoptic masking studied by McKee et al. (1994). In Figures 11-25, surface-to-boundary feedback does not change the result determined by the other model mechanisms. See text for an explanation.

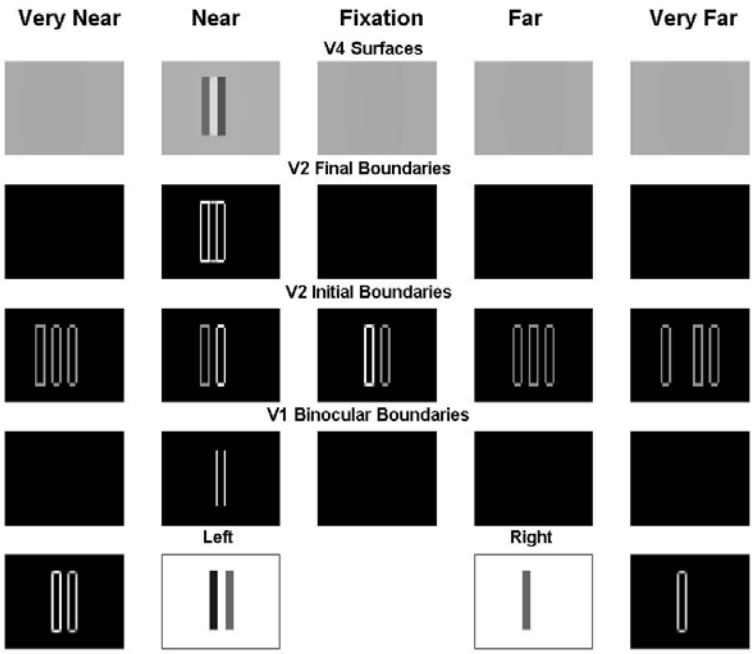


(a)

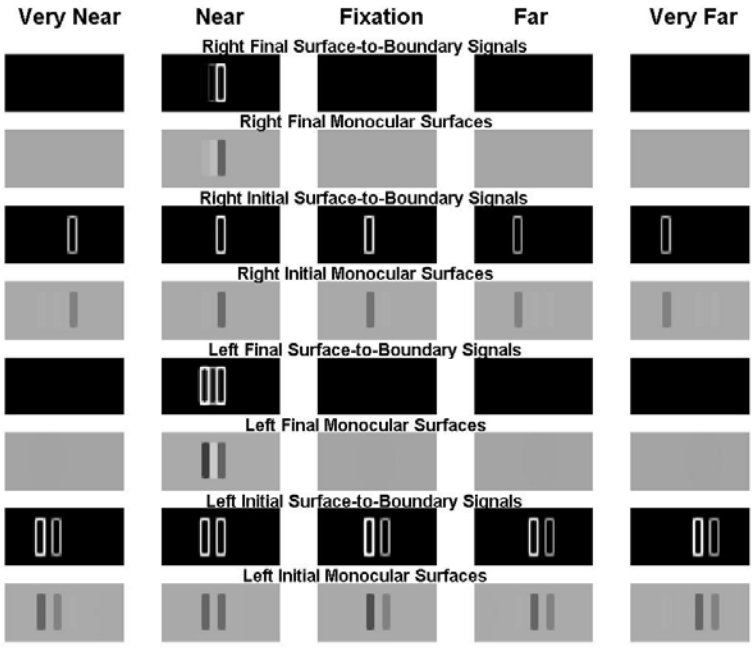


(b)

Figure 12. Simulation of the release from dichoptic masking studied by McKee et al. (1995).

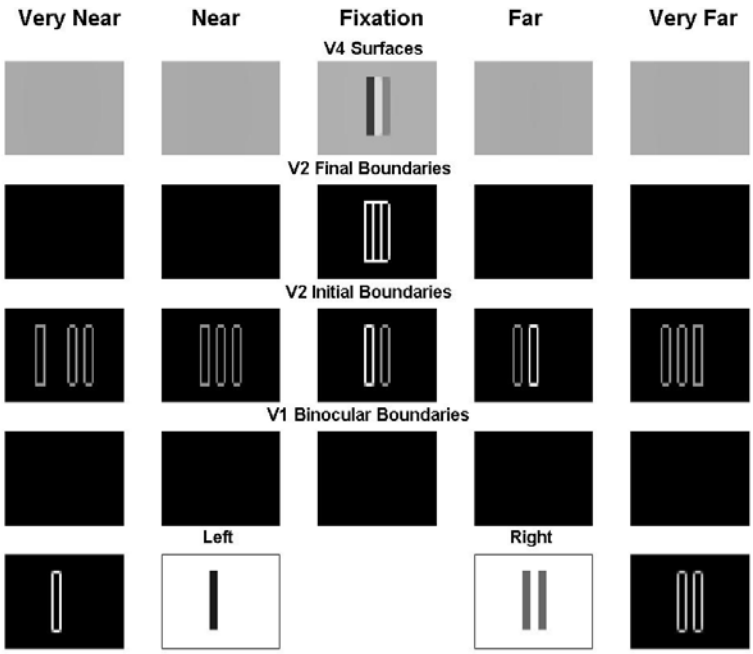


(a)

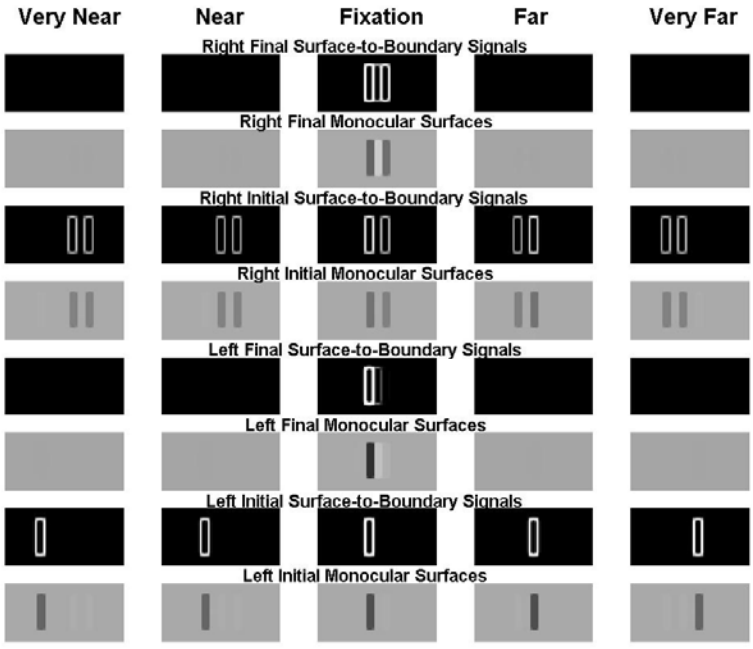


(b)

Figure 13. Simulation of a variation of the release from dichoptic masking studied by McKee et al. (1995).

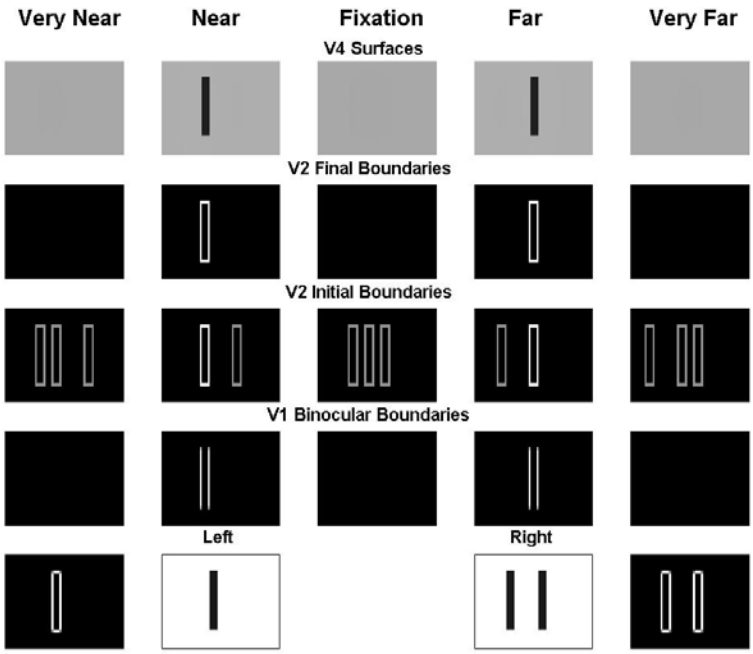


(a)

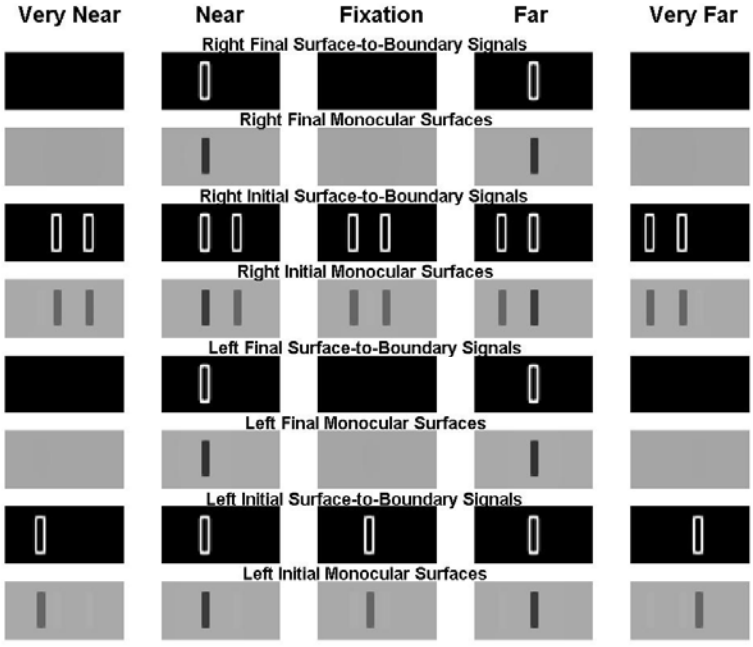


(b)

Figure 14. Simulation of the return to dichoptic masking studied by McKee et al. (1995).



(a)



(b)

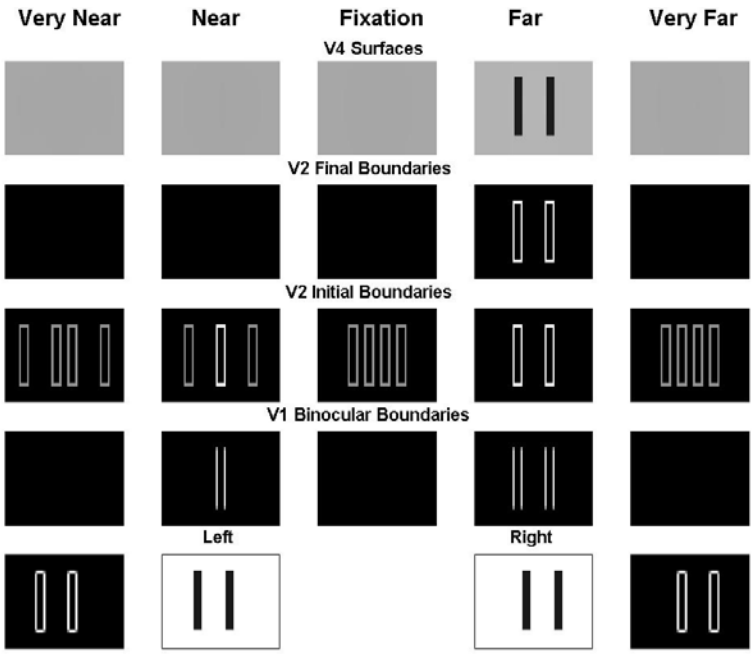
Figure 15. Simulation of dichoptic masking in Panum's limiting case studied by McKee et al. (1995).

simultaneously fused with the two right input bar and so masks them both equally as reported experimentally by McKee et al. (1995).

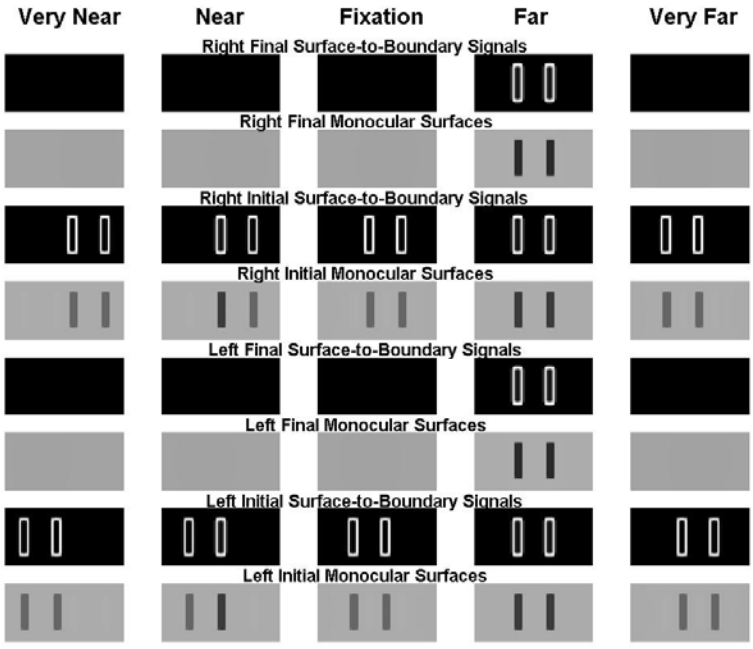
3.3.2. Contrast variations of the correspondence problem. In the control experiment by Smallman and McKee (1995), each eye was presented with two bars, all four bars having the same high contrast. Subjects reported seeing two bars, both in the far depth plane. The model simulation is shown in Figure 16. The enhanced model explains this case as well as the Grossberg and Howe (2003) model since the relative strength of boundaries is not undermined by surface-to-boundary feedback. The explanation is as follows: The vertical edges of the two bars are binocularly fused in the far depth plane in V1, as shown by the fourth plot of the second row. In addition, there is a false match in the near depth plane of V1, shown in the second plot of this row, which is caused by the inappropriate fusion of the right bar of the left input with the left bar of the right input. As usual, the monocular boundaries are added to all depth planes in V2 layer 4 along their respective monocular lines-of-sight, as shown in the third row of this figure. The binocular boundaries are also added to V2 layer 4 in their own depth planes, coinciding with the middle bar representation in the second plot and both bar representations of the fourth plot. All boundaries in V2 layer 4 then input into V2 layer 2/3, which controls the monocular surface filling-in within the V2 thin stripes. Those boundaries that receive binocular input, being stronger, initiate stronger surface filling-in which then causes stronger surface-to-boundary feedback signals to V2 layer 4. Therefore, surface-to-boundary feedback does not undermine the relative strength of V2 boundaries. The stronger boundaries quickly eliminate all other weaker vertical boundaries receiving only monocular input via disparity filter line-of-sight inhibition. Because they share monocular inputs, the two sets of vertical boundaries in the fourth plot, both of which receive binocular input, cooperate via the disparity filter to inhibit the vertical boundaries of the middle bar representation of the second plot, which also receive binocular input. The final V2 boundary representations are shown in the fourth row of Figure 16a. The model correctly predicts that subjects see both bars in the far depth plane.

In a more complicated version, either eye was presented with three identical bars with the same high contrast. The simulation is shown in Figure 17. The explanation is similar and hereby omitted.

In the first contrast variation experiment by Smallman and McKee (1995), either eye was presented with two bars, but the left bar of the left input had a much lower contrast than the other three bars. Subjects reported perceiving the lower contrast bar to lie in the fixation plane while also perceiving two high contrast bars, the left lying in a near depth plane and the right lying in a far depth plane. The explanation is similar to that of the Grossberg and Howe (2003) model since the relative strength of boundaries is still not undermined by surface-to-boundary feedback. A detailed explanation for this case will nonetheless be given because it shows that the bias in favor of the fixation plane is needed. The explanation is as follows: Because of its contrast difference, the left bar of the left input cannot be binocularly fused with either of the bars of the right input. Instead, the right bar of the left input binocularly matches with both bars of the right input in the near and far depth planes in V1, respectively, as shown by the second and fourth plots of the second row of Figure 18a. As usual, the monocular boundary representations are added to all depth planes in V2 layer 4 along their respective lines-of-sight, as shown by the plots in the third row of this figure. The V1 binocular boundary representations are also added to V2 layer 4, coinciding with the middle bar representation of the second plot and the right bar representation of the fourth plot. All boundaries in V2 layer 4 then input into V2 layer 2/3. The two sets of boundaries which receive binocular input from V1, being stronger, create stronger monocular

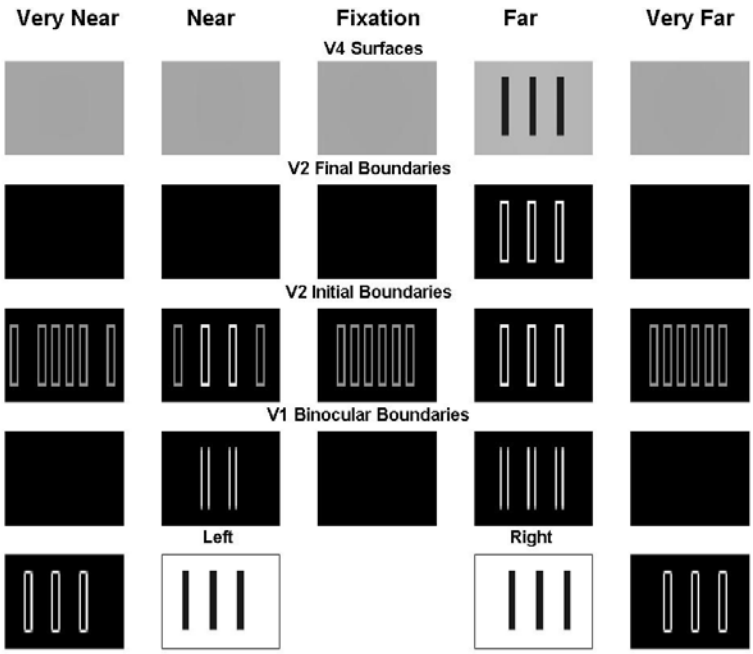


(a)

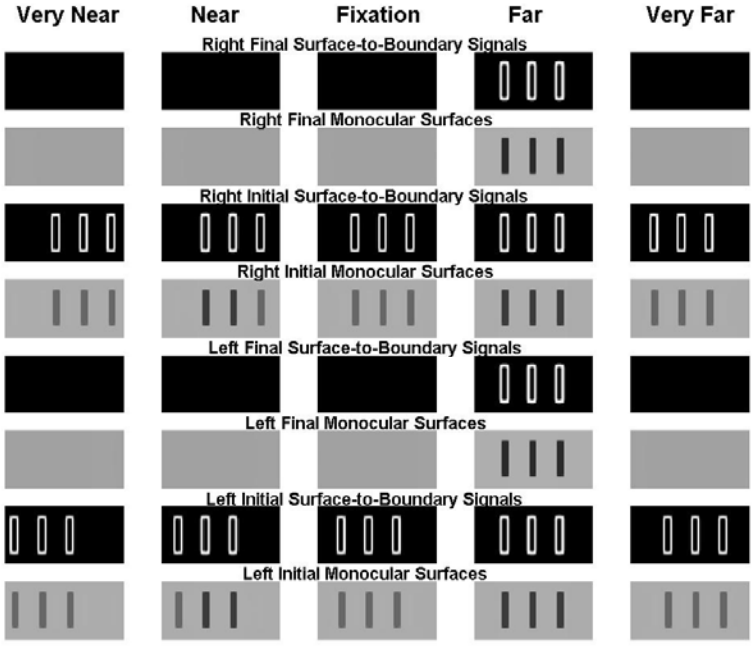


(b)

Figure 16. Simulation of the control experiment of the correspondence problem studied by Smallman and McKee (1995).

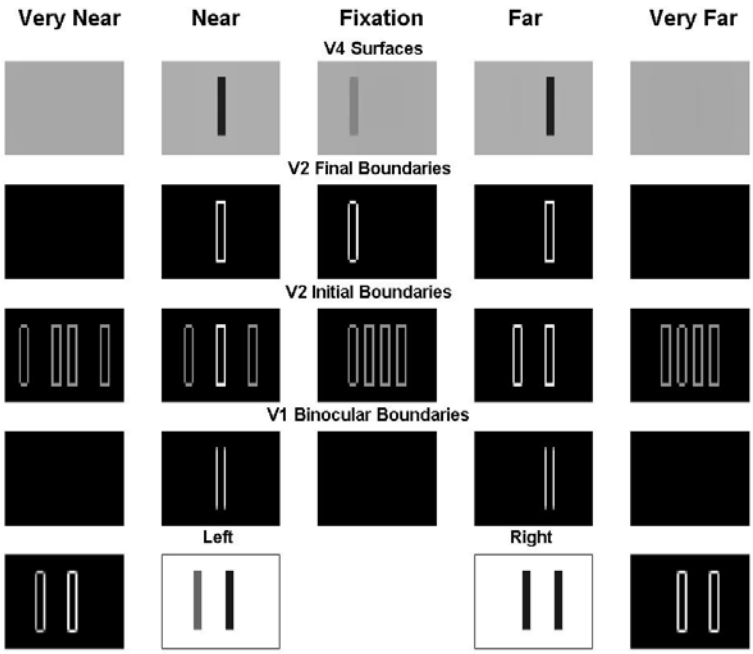


(a)

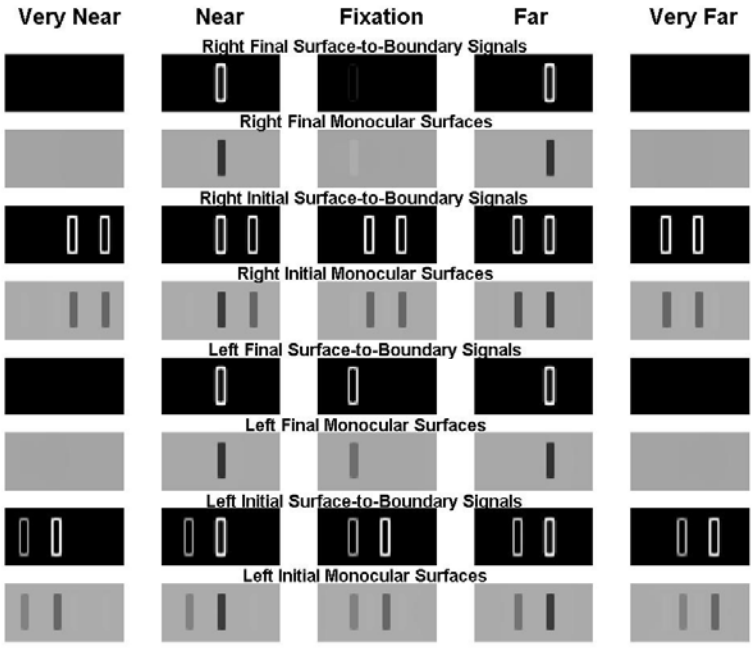


(b)

Figure 17. Simulation of a more complicated variation of Fig.16.

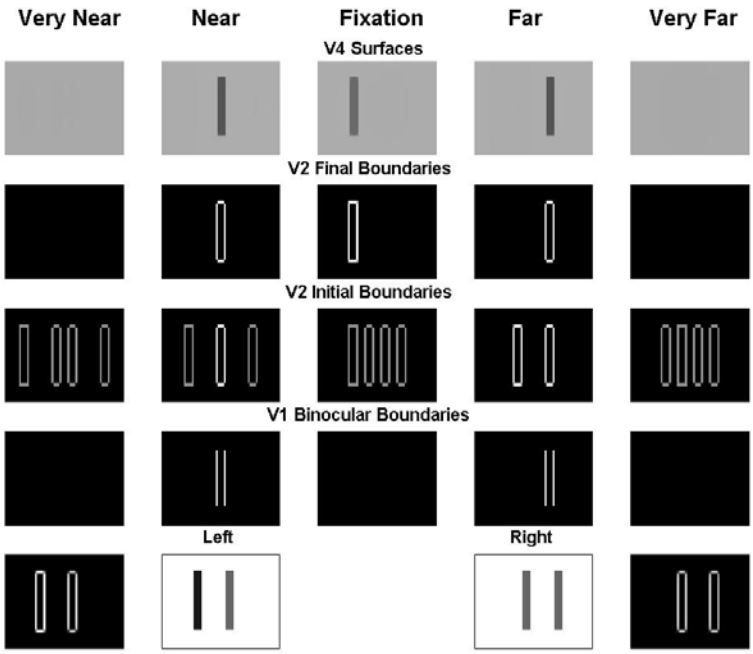


(a)

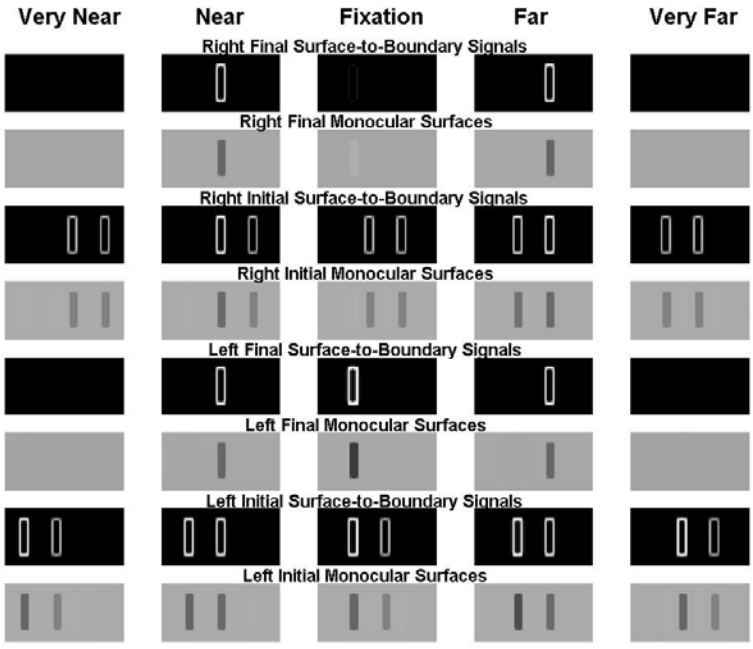


(b)

Figure 18. Simulation of a contrast variation of the correspondence problem reported by Smallman and McKee (1995).



(a)



(b)

Figure 19. Simulation of another contrast variation of the correspondence problem reported by Smallman and McKee (1995).

surfaces in the V2 thin stripes via filling-in, which then send stronger surface-to-boundary feedback signals to these stronger V2 boundaries. Therefore, surface-to-boundary feedback does not undermine the relative strength of V2 boundaries. The two sets of boundaries which receive binocular input from V1 remains stronger, which then eliminate those weaker vertical boundary representations that share their lines-of-sight via disparity filter line-of-sight inhibition.

However, except in the far depth plane, where the lower contrast bar coincides with the left high contrast bar of the right input, they do not eliminate the vertical boundary representations of the lower contrast bar in all other depth planes because these boundaries do not share any of their lines-of-sight. Here is where the bias to the fixation plane is relevant. Since the fixation plane is favored, the vertical boundary representation of the lower contrast bar in the fixation plane eliminates its boundary representations in other depth planes via disparity filter line-of-sight inhibition. The final V2 boundary representations are shown in the plots of the fourth row in Figure 18a. As usual, V4 fills in those regions that are completely enclosed by connected boundaries, as shown in the top row.

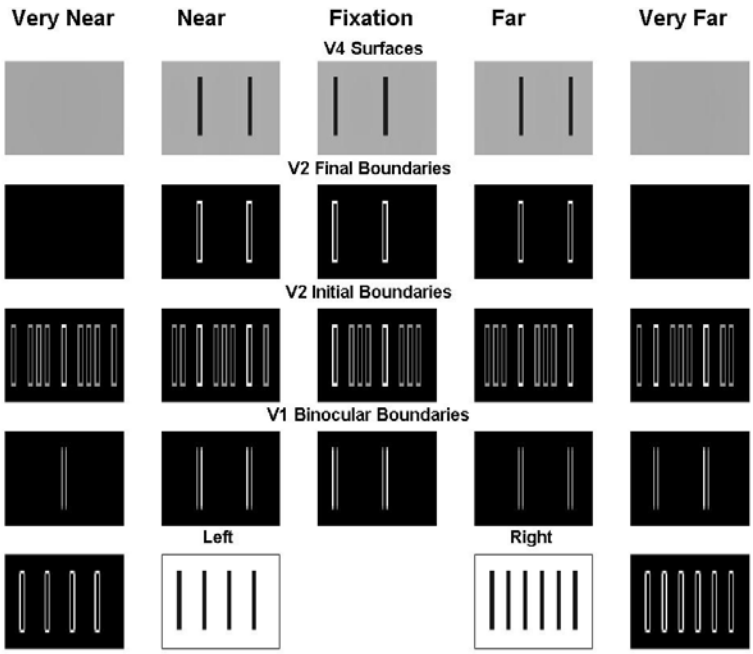
In the second contrast variation experiment, Smallman and McKee (1995) studied the inverse situation where the left bar of the left input had a much higher contrast than the other three bars, which all had the same contrast. The model simulation is shown in Figure 19. The explanation is similar and therefore omitted.

3.3.3. The Venetian blind effect. A Venetian blind stereogram consists of two gratings, a low frequency one that is presented to the left eye, and a high frequency one presented to the right (Howard and Rogers, 1995). When fused, the frequency of the gratings are such that every second bar of the left grating is in retinal correspondence with every third bar of the right grating. The model simulation is shown in Figure 20, which shows the perceived Venetian blind effect in the fifth row of Figure 20a. As before, surface-to-boundary feedback does not undermine the relative strength of V2 boundaries. The enhanced model therefore explains this case as well as Grossberg and Howe (2003), and thus the detailed explanation is omitted.

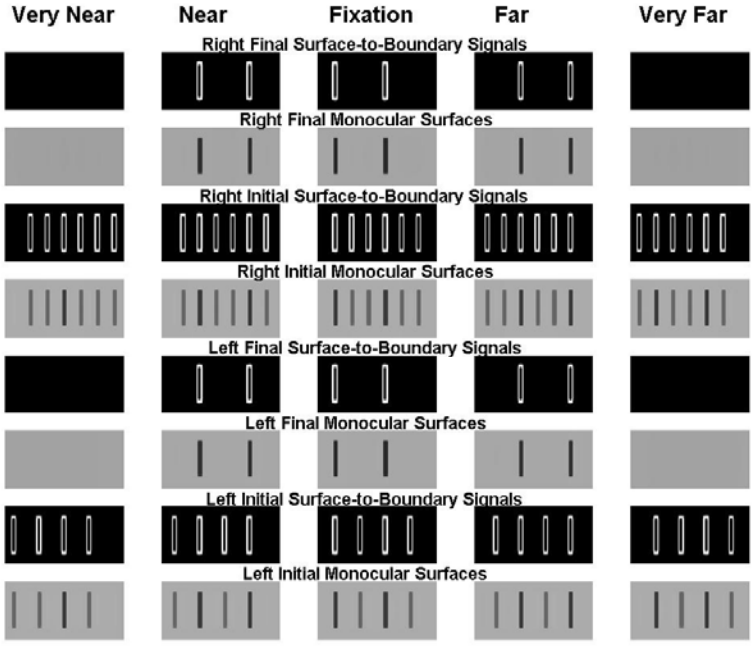
3.3.4. Experiments of Gillam et al. (1999). As discussed in Section 3.1, the enhanced model correctly predict variations of da Vinci stereopsis in Nakayama and shimojo (1990). Now we consider the da Vinci stereopsis experiments in Gillam et al. (1999). These experiments were simulated in Grossberg and Howe (2003), but the enhanced model gives a more consistent explanation.

In Figure 21, the right eye sees two thin bars and the left eye a single thick bar. Subjects report seeing two thin bars, the left in the near disparity plane and other in the far disparity plane. The model explanation is as follows. The left edge of the thick bar fuses with the left edge of the left thin bar to appear in a near disparity plane in V1, represented by the second plot of the second row, while the right edge of the thick bar fuses with the right edge of the right thin bar to appear in a far disparity plane in V1, represented by the fourth plot of this row, since in both cases these edges have the same contrast. The two other vertical edges of the thin bars of the right input are registered only monocularly because they cannot be matched to either of the edges of the left input.

As usual, the V1 monocular boundary representations are added to all depth planes in V2 layer 4 along their respective lines-of-sight. The V1 binocular boundary representations are also added to V2 layer 4, overlapping with the leftmost vertical boundary in the second plot and the rightmost vertical boundary in the fourth plot. All boundaries in V2 layer 4 then input into V2 layer 2/3 which acts as disparity filter. The boundaries in V2 layer 2/3 also control the formation of monocular surfaces in the V2 thin stripes, which generate surface-to-boundary feedback

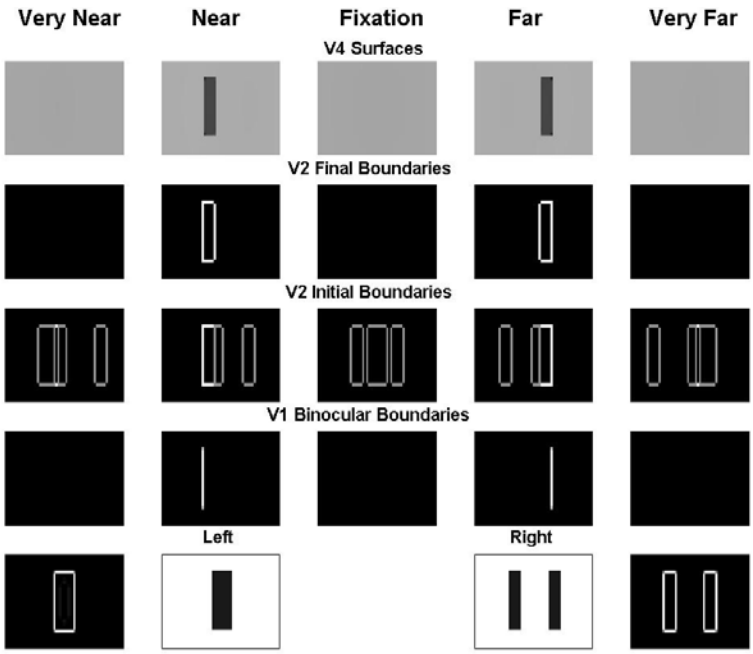


(a)

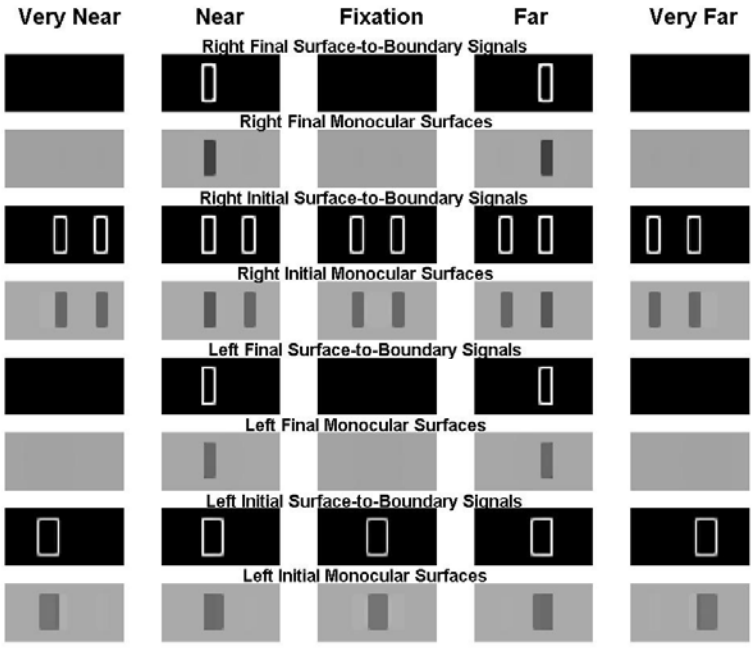


(b)

Figure 20. Simulation of the Venetian blind effect in Howard and Rogers (1995).

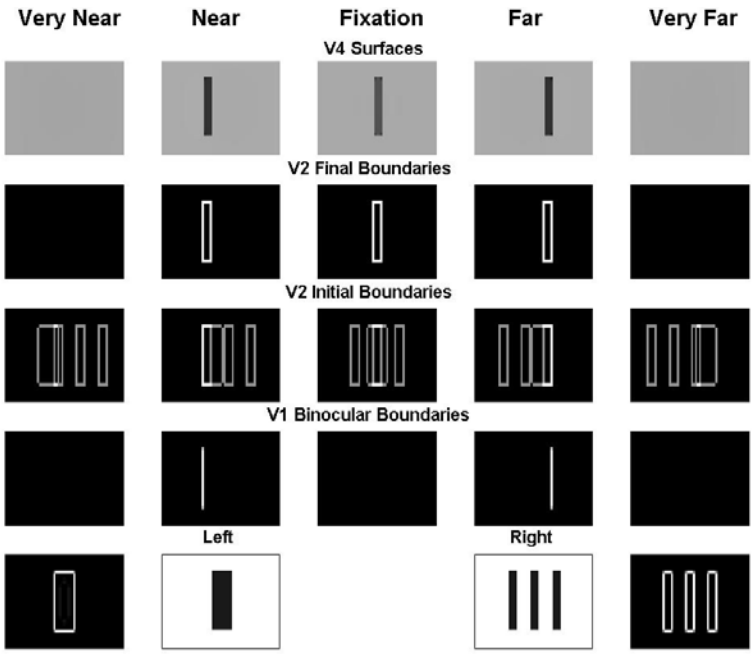


(a)

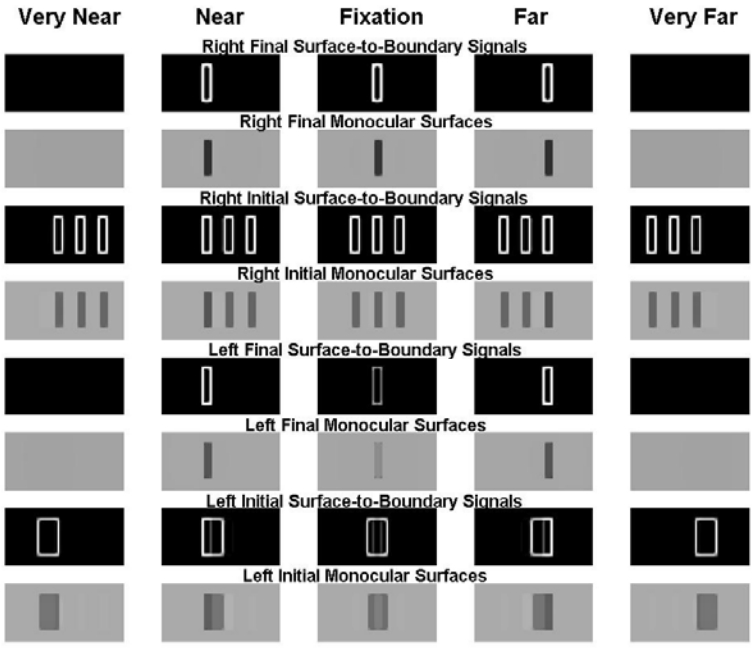


(b)

Figure 21. Simulation of the da Vinci stereopsis studied by Gillam et al. (1999).



(a)



(b)

Figure 22. Simulation of a variation of the original da Vinci stereopsis of Fig. 21.

signals. Initially all boundaries receive such signals. The stronger vertical boundaries, which receive binocular inputs from V1 − namely, the leftmost vertical boundary in the second plot and the rightmost vertical boundary in the fourth plot − eliminate all the other weaker vertical boundaries that share any of their lines-of-sight via the line-of-sight inhibition of the disparity filter. Although they do not inhibit those vertical boundary representations originating from the two monocularly viewed edges of the right input, because these vertical boundaries do not share any of their lines-of-sight, the vertical boundary representations of monocularly viewed edges in the near and far depth planes may be defeated by their corresponding representations in the fixation plane, which is biased.

This does not happen due to surface-to-boundary feedback. The reason is as follows. The boundary representations of binocularly viewed edges of the right input in the fixation plane are quickly eliminated by their stronger corresponding representations in the near and far planes. Once this happens, the connected boundaries for the two thin bars of the right input in the fixation plane are both destroyed, and hereby the corresponding monocular surfaces cannot be filled in. As a result, the boundary representations in the fixation plane of two monocularly viewed edges of the right input do not receive surface-to-boundary feedback signals and hereby are suppressed. The boundary representations of two monocularly viewed edges of the right input in the near and far depth planes thus survive. The final V2 boundary representations are shown in the fourth row. As usual, V4 fills-in surfaces in those regions that are completely enclosed by connected boundaries, resulting in the percept of a thin near bar and a thin far bar, as reported experimentally by Gillam et al. (1999).

In the previous display, at least one edge of each region could be binocularly fused. In contrast, Figure 22 depicts a variant of the original Gillam et al. display. Here, the left eye sees a single bar while the right eye sees three separate bars. The middle bar of the right eye stimulus cannot be binocularly fused with any vertical edge of the left bar.

The model explanation of this percept is as follows. The left edge of the bar of the left input again fuses with the left edge of the leftmost bar of the right input to form a binocular boundary in the second plot of the second row of Figure 22a. The right edge of the bar of the left input again fuses with the right edge of the rightmost bar of the right input to form a binocular boundary in the fourth plot of the second row. As usual, the monocular boundaries are added to V2 layer 4 along their respective lines-of-sight, as shown by the third row. The binocular V1 boundaries are also added to V2 layer 4, overlapping with the first vertical boundary in the second plot and the last vertical boundary in the fourth plot of the third row. All boundary signals in V2 layer 4 then input into V2 layer 2/3. The boundary grouping in layer 2/3 controls monocular surface filling-in in the V2 thin stripes.

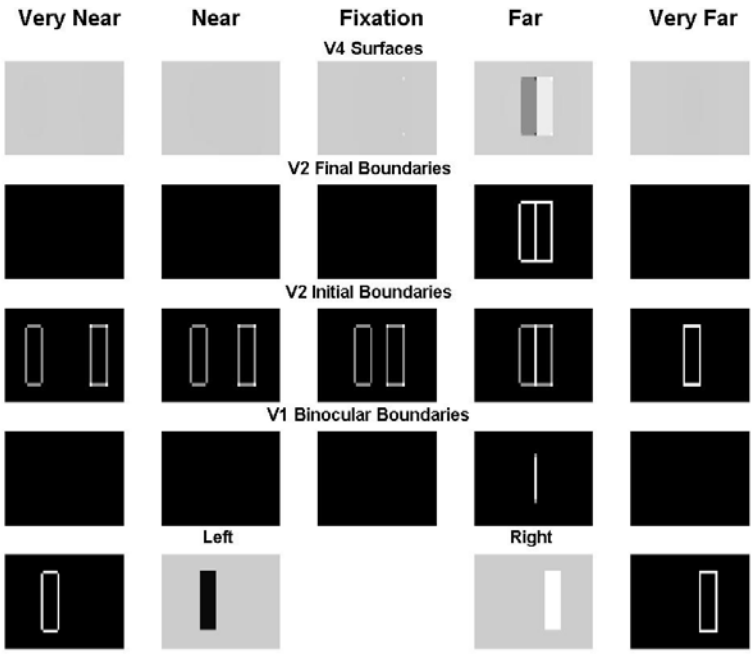
The successfully filled-in surfaces send surface-to-boundary feedback signals to V2 layer 4. Initially, all monocular surfaces are successfully filled-in, and hereby all boundaries in V2 layer 4 receive surface-to-boundary signals. The V2 vertical boundaries which receive binocular inputs from V1 − namely, the first vertical boundary in the second plot and the last vertical boundary in the fourth plot of the third row − being much stronger, eliminate all the other weaker vertical boundaries that share any of their lines-of-sight via the line-of-sight inhibition of the disparity filter. In particular, the first and last vertical boundaries in the fixation plane are eliminated. As a result, the monocular surface representations of the outer two thin bars of the right input associated with the fixation plane cannot be filled in. Therefore, the remaining boundaries of the two thin bars in the fixation plane are suppressed, since they do not receive surface-to-boundary signals any more, and hereby the corresponding vertical boundary

representations of monocularly viewed edges of the two thin bars in the near and far depth planes survive, even though the fixation plane is favored. The vertical boundary representations of the middle bar of the right input in the fixation plane win against their corresponding representations in other depth planes, since the fixation plane is favored. The final V2 boundaries are shown in the fourth row of Figure 22a. As usual, V4 fills-in surfaces in those regions that are completely enclosed by connected boundaries. The model therefore correctly predicts that three surfaces will be seen, each at a different depth, as reported experimentally by Gillam et al. (1999).

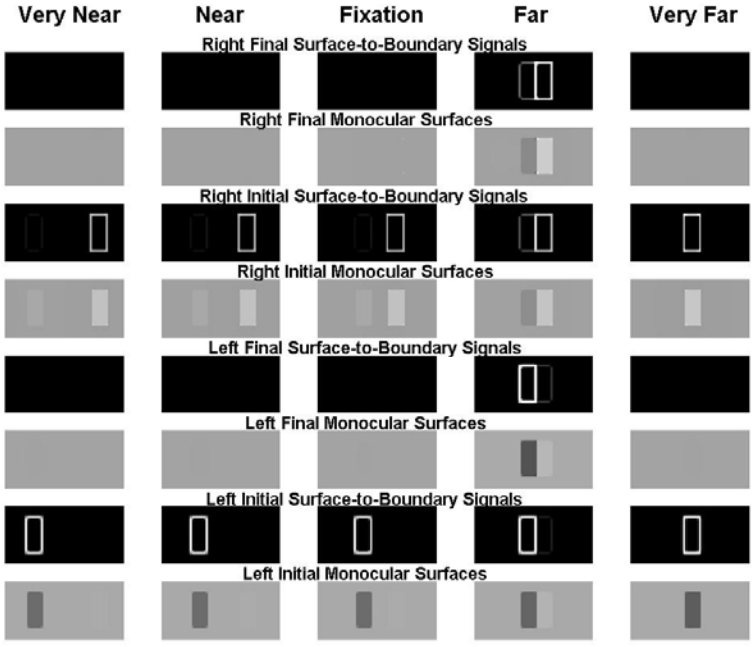
3.3.5. Stereopsis with opposite-contrast stimuli. A polarity-reversed stereogram studied by Grossberg and Howe (2003) is shown in the middle two plots in the bottom row of Figure 23a, where the left eye sees a black bar and the right eye a white bar both on the same gray background. Subjects reported perceiving a black bar abutting a white bar in a far depth plane. The enhanced model can give a better explanation of this percept than the Grossberg and Howe (2003) model. The initial stages of this explanation are similar to those of the Grossberg and Howe (2003) model. Since they have the same contrast polarity, the left edge of the white bar binocularly fuses with the right edge of the black bar to form a vertical boundary in the far depth plane of V1, as shown by the fourth plot of the second row. As usual, the monocular boundaries are added to V2 layer 4 along their respective lines-of-sight, as shown in the third row. The vertical binocular boundary in V1 is also added to V2, coinciding with the middle vertical boundary in the far depth plane, as shown in the fourth plot of the third row. All boundaries in V2 layer 4 then input into V2 layer 2/3, which controls monocular surface filling-in within the V2 thin stripes. The successfully filled-in surfaces then send surface-to-boundary feedback signals to V2 layer 4. Initially, all monocular surfaces are successfully filled-in and therefore all V2 boundaries receive surface-to-boundary feedback signals. The V2 boundary which receives input from the V1 binocular boundary − namely, the middle boundary in the fourth plot of the third row of Figure 23a − being much stronger, quickly eliminates the middle two vertical boundaries in all other depth planes via disparity filter line-of-sight inhibition. This destroys the connected boundary contours in all depth planes other than the far depth. As a result, the corresponding V2 monocular surfaces are no longer successfully filled-in. As a result, except for those in the far depth plane, all remaining boundaries in other depth planes are suppressed since they do not receive surface-to-boundary signals anymore. Therefore, even though the fixation plane is favored, the outer two vertical boundaries in the far depth plane can survive, as shown in the fourth plot of the fourth row of Figure 23a. As usual, V4 fills in those regions that are completely enclosed by connected boundaries, as shown in the top row. The model hereby makes the correct prediction.

A variation of the above polarity-reversed stereogram is shown in the middle two plots of the bottom row of Figure 24a, where the black bar and the white bar are in the corresponding retinal positions. Since the polarities are reversed, they cannot fuse in that position. Subjects correspondingly report no stable surface percepts (Grossberg and Howe, 2003). The model simulation is showed in Figure 24. Since surface-to-boundary feedback does not reduce the relative strength of the stronger boundaries, the model explanation is very similar to that in Grossberg and Howe (2003) model and is thus omitted.

Nakayama and Shimojo (1990) studied the polarity-reversed da Vinci stereopsis stimuli shown in the middle two plots in the bottom row of Figure 25a, where the binocularly viewed thick bar and monocularly viewed thin bar had opposite luminance polarities. The model simulation is shown in Figure 25.

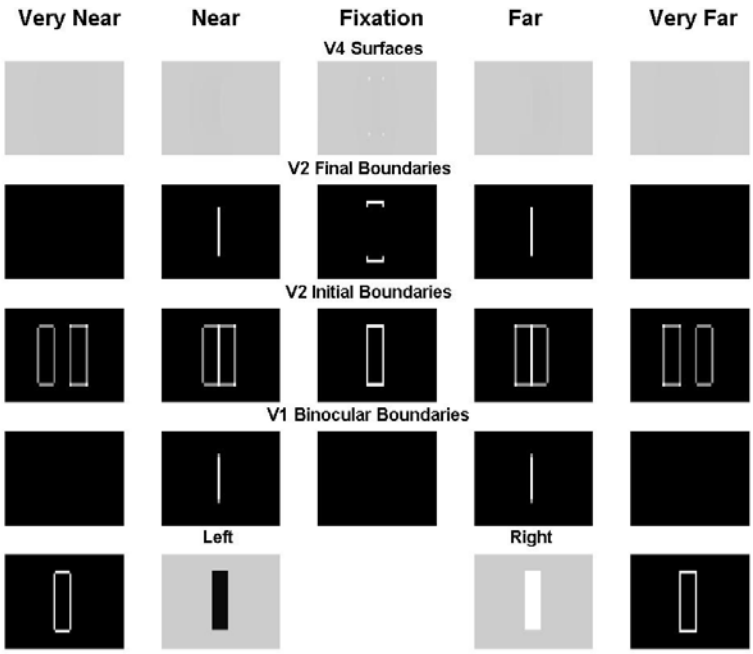


(a)

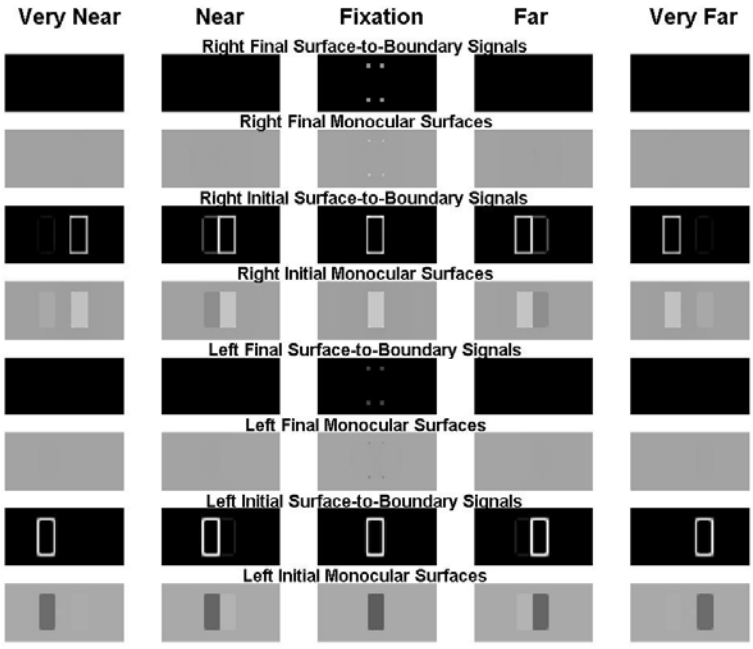


(b)

Figure 23. Simulation of stereopsis with a polarity-reversed sterogram.

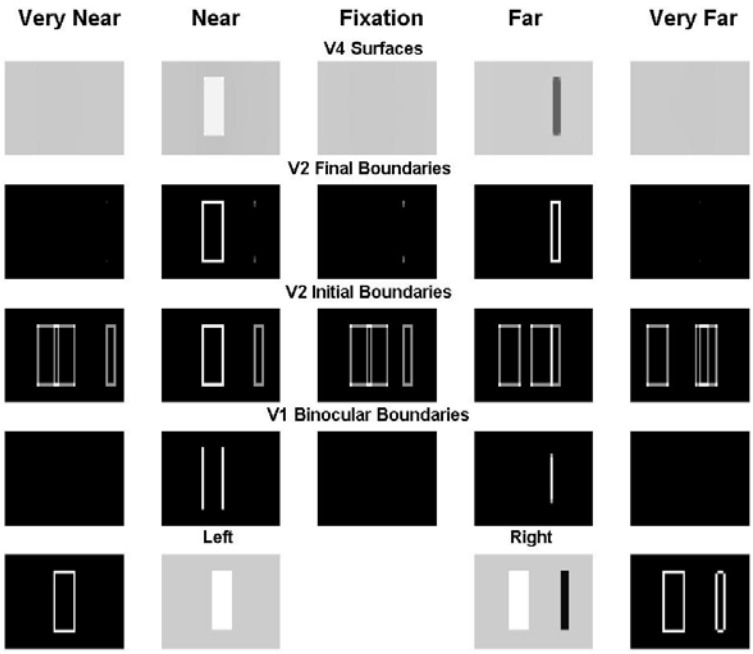


(a)

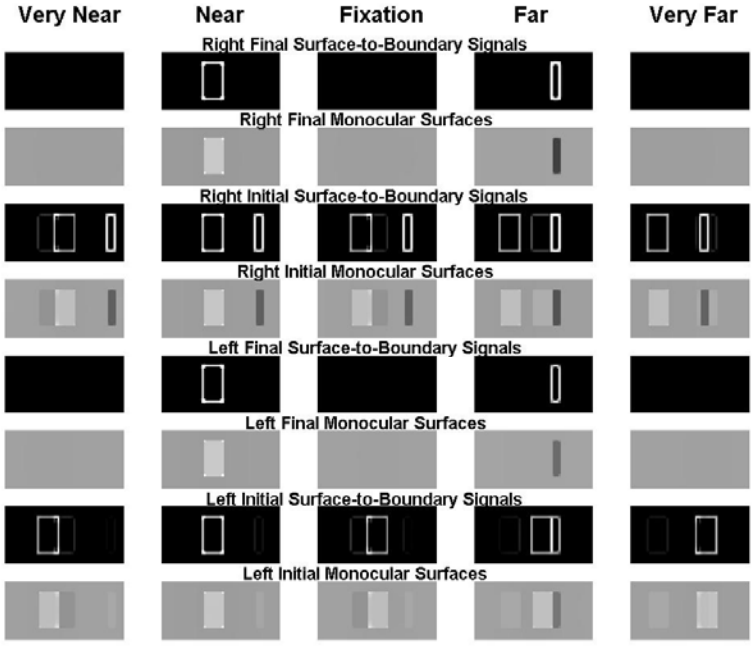


(b)

Figure 24. Simulation of a variation of Fig. 23.



(a)



(b)

Figure 25. Simulation of the polarity-reversed da Vinci stereopsis studied by Nakayama and Simojo (1990).

The explanation is the same as the original da Vinci stereopsis case in Figure 6, except here the right edge of thick bar of the left input fuses with left edge of thin bar since they have the same contrast polarity.

4. Discussion

The present article predicts how complementary boundary and surface processes work together to explain challenging psychophysical data about stereopsis and 3D planar surface perception. In so doing, it shows how mechanisms that have previously helped to explain data about 3D figure-ground separation, notably the separation in depth of occluding and occluded objects, also help to explain data about stereopsis per se.

To realize this goal, the article develops a laminar model of visual cortex that embodies two major new hypotheses: (1) *Grouping suppresses false matches*: The model predicts how long-range grouping processes in layer 2/3 of V2 include inhibitory interneurons which realize a disparity filter that eliminates false matches. In other words, the Gestalt grouping problem subsumes the binocular false match problem. The model predicts that the disparity filter consists of line-of-sight inhibitory interactions in the pale stripes of V2; see Figure 5a. A disparity filter with line-of-sight inhibition was first introduced into the modeling literature to simulate stereopsis data in Grossberg and McLoughlin (1997) and McLoughlin and Grossberg (1998), but without a laminar interpretation. These inhibitory interactions include a fixation plane advantage, which may develop due to expansion of cortical receptive fields at the fixation plane. The model predicts that disparity filter inhibition is part of a network of inhibitory interactions that also includes inhibitory interneurons, which are predicted to interact with long-range excitatory horizontal connections to achieve a *bipole* property during perceptual grouping (Grossberg, Mingolla, and Ross, 1997; Grossberg and Raizada, 2000). Grossberg and Williamson (2001) simulated how the inhibitory interneurons that subserve the bipole property may develop in response to visual inputs. Similar developmental rules may also control development of the line-of-sight inhibitory connections that form part of the disparity filter. Demonstrating this remains a task for future work. (2) *Surface-to-boundary feedback modulates grouping strength*: Reciprocal connections between the thin (surface) and pale (boundary) stripes in V2 have earlier been predicted to ensure that the boundaries and surfaces that emerge in conscious percepts are *consistent*, even though the rules whereby they are formed are *complementary* (Grossberg, 1994, 2000). These reciprocal interactions are, we would suggest, one reason why the hypothesis that there are functionally separate cortical processing streams is continually challenged. Our results clarify how these complementary intra-stream rules may seem to be blended, at least as measured in neurophysiological recordings, due to reciprocal interactions between the streams. The simulations herein suggest that surface-to-boundary feedback makes unnecessary the cyclopean inhibition, or inhibition across depth but within position, that was needed by Grossberg and Howe (2003) to explain some stereopsis data. Hypothesis (1) is new. Hypothesis (2) has been used before to explain and simulate a wide range of data about 3D figure-ground perception (Grossberg, 1994, 1997; Kelly and Grossberg, 2000). It was also used to simulate one case of da Vinci stereopsis by Grossberg and McLoughlin (1997). Here it is shown how surface-to-boundary feedback helps to explain a wider range of daVinci stereopsis cases than the Grossberg and Howe (2003) model could handle, and does so without the need for cyclopean inhibition within the disparity filter.

One of the main predicted properties of surface-to-boundary feedback is that only *connected* boundaries, which can contain filling-in within the surface stream, can generate

feedback signals to the boundary stream at the contours of the filled-in surfaces, and that horizontal boundaries are added to all depth planes for this to happen (Figure 4). Yazdanbakhsh and Watanabe (2004) have successfully tested this prediction.

The combination of FACADE figure-ground mechanisms and 3D LAMINART stereopsis mechanisms that were unified in this article have made it possible to explain an even wider range of data about 3D vision and figure-ground perception than was previously possible. These include stratification, transparency, and 3D neon color spreading (Grossberg and Yazdanbakhsh, 2003a, 2003b, 2005), 3D percepts of curved and slanted 2D pictures and 3D scenes, including bistable percepts such as the Necker cube and binocular rivalry (Grossberg and Swaminathan, 2004; Yazdanbakhsh and Grossberg, 2005), and the perception of sparse and dense textured Julesz stereograms (Fang and Grossberg, 2004).

The work on transparency makes detailed predictions about how multiple parts of visual cortex work together to determine when surfaces appear opaque or transparent. In particular, the Grossberg and Yazdanbakhsh (2005) transparency simulations utilize the predicted mechanism that boundaries of occluded surfaces are broken when an occluding surface appears in front. Such a boundary gap predicts that contrasts can flow, of fill-in, between the surface regions that are separated by the broken boundary, thereby leading to a change of perceived brightness. Tse (2005) has demonstrated changes of perceived brightness that are consistent with this prediction during percepts of bistable transparency. The model predicts that color mixing should be perceivable across such a boundary gap in a bistable transparency display that is suitably constructed, say by using watercolor illusion stimuli (Pinna, Brelstaff, and Spillmann, 2001; Pinna and Grossberg, 2005). Multiple modeling and data constraints suggest that surface filling-in occurs at more than one stage of cortical processing between V1, V2, and V4 (Grossberg, 1994; Grossberg and Hong, 2004; Kelly and Grossberg, 2000). It remains to definitely test whether the predicted interactions between cells in different thin stripes of V2 support one level of this predicted surface filling-in process.

The proposed explanations and simulations of slanted and curved surface perception, notably simulations of Necker cube bistability and binocular rivalry, consistently extend the mechanisms that are used to simulate the data herein. In particular, Grossberg and Swaminathan (2004) showed how bipole cells, angle cells, and disparity-gradient cells can all develop from the same developmental growth rules in response to different statistics of visual scenes. Grossberg and Sqaminathan (2004) summarize psychophysical and neurobiological evidence that supports the existence of these cell types. Angle and disparity-gradient cells are predicted to enable boundary groupings to form across multiple depths. For example, in a 2D picture of a 3D cube, a line that is neither horizontal nor vertical can often be part of a surface whose depth changes from near-to-far or from far-to-near, depending on where it is attached to the cube. The same line can be part of a flat 2D surface as well. Contextual interactions between angle cells, disparity-gradient cells, and bipole cells are predicted to determine which depth relations (flat, near-to-far, far-to-near) are perceived in the cube percept. In the Necker cube simulations, boundary gaps form in the weaker of two intersecting boundaries. As in the case of transparency, these gaps enable one surface to be seen in front of another, until the percept switches. The Tse (2005) experiment supports this explanation of bistable Necker cube percepts as well as those of bistable transparency.

In addition to these expanding the cell types involved in perceptual grouping, the work on bistable percepts predicts that habituative transmitter gates (Grossberg, 1968, 1969, 1980;

Francis, Grossberg, and Mingolla, 1994), also called depressing synapses (Abbott et al., 1997), help to weaken an active perceptual representation and thereby enable a competing perceptual interpretation to supplant it. Earlier work predicted that these habituative transmitters play a basic role in controlling the development of visual cortex (Grossberg, 1976; Grossberg and Seitz, 2003) and in controlling the form-sensitive reset of boundary groupings in response to changing scenes (Francis, Grossberg, and Mingolla, 1994). Necker cubes and binocular rivalry stimuli just happen to confront these mechanisms with an approximately balanced set of two or more perceptual representations. The fact that the habituative transmitters exist within recurrent cooperative and competitive circuits that are used in perceptual filtering, grouping, and attention is predicted to lead to the perceived switches in the two types of percepts.

The work on Julesz stereograms provides the first simulations of how perceived lightnesses are lifted, via surface capture, to their correct surfaces in depth when, as in a dense stereogram, the correspondence problem is a serious challenge. This work also shows how surfaces can be perceived at different depths even if their inducers are sparse. Both of these explanations rely on the types of 3D boundary groupings that are simulated here, as well as upon how these 3D boundaries can selectively capture lightnesses on surfaces that are seen at different depths. One key stage of this depthful-surface capture process is predicted to occur in the thin stripes of cortical area V2; see Figure 5 and Grossberg (1994, 1997) for further discussion of the surface capture process. These stereogram simulations have brought the 3D LAMINART model to the threshold of being able to process natural 3D scenes.

In all of these cases, detailed properties of the neurophysiology and anatomy of visual cortex are functionally linked to specific properties of conscious percepts. Given the range of phenomena that now have explanations, new types of experiments have been designed to further test the predicted processes. Some of the successfully tested predictions are summarized above. Every model neurophysiological property of an identified cell type, anatomical connection between identified cell types, and perceptual property that emerges from rigorously characterized model interactions among these neurobiologically identified elements constitutes a prediction that can be tested by multiple experimental methods using several different types of visual stimuli. Additional tests of these interdisciplinary predictions are eagerly awaited.

Appendix A. Model Equations

This section describes the model equations. Each neuron is typically modeled as a single voltage compartment in which the membrane potential, *v(t)*, is given by

$$C_m \frac{dv(t)}{dt} = (E_{leak} - v(t))g_{leak} + (E_{excit} - v(t))g_{excit}(t) + (E_{inhib} - v(t))g_{inhib}(t), \quad \text{(A1)}$$

where parameters *E* denote reversal potentials, g_{leak} is a constant leakage conductance, and the time-varying conductances $g_{excit}(t)$ and $g_{inhib}(t)$ represent the total excitatory and inhibitory inputs to the cell (Grossberg, 1973; Hodgkin, 1964). By setting the capacitance term $C_m = 1$, the leakage conductance $g_{leak} = A$, and the reversal potentials $E_{excit} = B$, $E_{inhib} = -C$, and $E_{leak} = 0$, the membrane equation (A1) can be rewritten in the form,

$$\frac{dv}{dt} = -Av + (B - v)g_{excit} - (C + v)g_{inhib}, \quad \text{(A2)}$$

where *A* is a constant decay rate, *B* is the maximum membrane potential, *C* is the minimum membrane potential, g_{excit} is the total excitatory input, and g_{inhib} is the total inhibitory input. In the following simulations, a single set of parameters was used. For simplicity, the same parameter symbol (e.g., α) in different equations may be used, but it does not always represent the same value. Its value in each equation is given below. Each eye's stimulus was presented on a grid 30 units high and 60 units wide except Figure 17 which was presented on a 30 by 70 grid and Figure 20 which was presented on a 30 by 115 grid. In all simulations, the background had a luminance value, in arbitrary units, of 2. In Figures 6-22, the light gray bars (if any) had a luminance of 0.4 and the dark gray bars 0.1. In Figures 23-25, white bars were represented by a luminance of 40, and black by 0.1. Simulations were performed using the *Matlab* software package.

LGN. The LGN cells obey membrane equations that receive input from the retina and are assumed to have circularly symmetric on-center, off-surround receptive fields. When these fields are approximately balanced, the network discounts the illuminant and contrast-normalizes its cell responses (Grossberg and Todorović, 1988). The LGN cell membrane potentials, $x_{ij}^{L/R}$, obey the following differential equation:

$$\frac{dx_{ij}^{L/R}}{dt} = -\alpha x_{ij}^{L/R} + \left(\beta - x_{ij}^{L/R}\right)I_{ij}^{L/R} - x_{ij}^{L/R}\sum_{p\neq i, q\neq j} G_{pqij} I_{pq}^{L/R}, \quad \text{(A3)}$$

where L/R designates that the cell belongs to the left or right monocular pathway, indices i and j denote the position of the input on the retina, α is a constant (10^{-5}) that represents the rate of decay of the cell membrane potential, β is a constant (9.9) that represents the maximum membrane potential, $I_{ij}^{L/R}$ is the luminance of the left or right retinal image that represents the excitatory on-center, and G_{pqij} is a Gaussian kernel that represents the inhibitory off-surround:

$$G_{pqij} = \exp\left(-\frac{(p-i)^2 + (q-j)^2}{2\sigma^2}\right), \quad \text{(A4)}$$

where σ represents the size of the kernel (1.5). The steady-state cell membrane potentials of these cells are given by:

$$x_{ij}^{L/R} = \frac{\beta I_{ij}^{L/R}}{\alpha + \sum_{p,q} G_{pqij} I_{pq}^{L/R}}. \tag{A5}$$

The steady-state equation (A5) was used in the simulations.

V1 layer 4 simple cells. All cells in V1 layer 4 are modeled as monocular simple cells that are sensitive to either dark-light or light-dark contrast polarity, but not both, depending on their receptive field structure. At steady-state, the membrane potentials, $s_{ijk}^{L/R,+}$, of simple cells that respond to dark-light contrast polarity are given by:

$$s_{ijk}^{L/R,+} = \sum_{p,q} \mathrm{K}_{pqk} \left[x_{i+p,j+q}^{L/R} \right]^+. \tag{A6}$$

where the superscript L/R,+ means L+ or R+. L denotes left monocular, R denotes right monocular, + indicates that the simple cell responds to dark-light contrast polarity, and *k* denotes orientation. Two orientations (vertical (*k=1*) and horizontal (*k=2*)) were used in these simulations, the threshold linear function $[x]^+ = \max(x,0)$, and K_{pqk} is a Gabor function representing the simple cell receptive field kernel:

$$\mathrm{K}_{pqk} = \phi \sin\left(\frac{2\pi(r-0.5)}{\tau} \right) \exp\left[-\frac{1}{2} \left(\frac{(p-0.5)^2}{\sigma_p^2} + \frac{(q-0.5)^2}{\sigma_q^2} \right) \right], \tag{A7}$$

where ϕ , τ , σ_p , σ_q are constants (4.4, 3π, 0.6, 0.6) representing the amplitude and dimensions of this kernel; $r = p$ for cells that respond to vertical boundaries; and $r = q$ for those that respond to horizontal boundaries.

The cell membrane potentials of the simple cells with light-dark contrast polarity are the inverse of the previous cell membrane potentials:

$$s_{ijk}^{L/R,-} = -s_{ijk}^{L/R,+} = -\sum_{p,q} \mathrm{K}_{pqk} \left[x_{i+p,j+q}^{L/R} \right]^+. \tag{A8}$$

V1 Layer 3B monocular simple cells. At steady-state the membrane potentials, $b_{ijk}^{L/R,+/-}$, of the layer 3B monocular simple cells are given by:

$$b_{ijk}^{L/R,+/-} = 2\left[s_{ijk}^{L/R,+/-} \right]^+. \tag{A9}$$

where the multiplicative factor of 2 compensates for the fact that the monocular simple cells receive inputs from only one eye whereas the binocular simple cells, discussed in the next section, receive input from both eyes.

V1 Layer 3B binocular simple cells. The layer 3B binocular simple cells receive excitatory input from layer 4 and inhibitory input from the layer 3B inhibitory interneurons that correspond to the same position and disparity. The membrane potentials, $b_{ijkd}^{B,+/-}$, of layer 3B simple cells obey the equation:

$$\begin{aligned} \frac{d}{dt} b_{ijkd}^{B,+/-} = & -\gamma_1 b_{ijkd}^{B,+/-} + (1 - b_{ijkd}^{B,+/-}) \left(\left[s_{(i+s)jk}^{L,+/-} - \theta \right]^+ + \left[s_{(i-s)jk}^{R,+/-} - \theta \right]^+ \right) \\ & - \alpha \left(\left[q_{ijkd}^{L,+/-} \right]^+ + \left[q_{ijkd}^{L,-/+} \right]^+ + \left[q_{ijkd}^{R,+/-} \right]^+ + \left[q_{ijkd}^{R,-/+} \right]^+ \right), \end{aligned} \tag{A10}$$

where γ_1, α and θ are constants (0.1, 7.2, 0.4) representing the rate of decay of the membrane potential, the strength of the inhibition and the signal threshold, $q_{ijkd}^{L/R,+/-}$ are the membrane potentials of inhibitory interneurons in layer 3B, d is the disparity to which the model neuron is tuned and index s is the positional shift between left and right eye inputs that depends on the disparity and is defined in Table 1. Index k is the orientation. Here k is vertical since vertical left and right eye boundaries are matched.

Table 1. The allelotropic shift (s) is the amount that the left and right monocular contours must be displaced to form a single fused binocular contour. It depends on the disparity. It is zero for matches in the fixation plane because these matches are between contours at retinal correspondence. Figure 3 illustrates the allelotropic shift and shows that a left monocular contour needs to be shifted more to the right for matches that are further from the observer, whereas a right monocular contour needs to be shifted in the opposite direction.

Disparity (d)	**V. Near disparity**	**Near disparity**	**Zero disparity**	**Far disparity**	**V. Far disparity**
Allelotropic shift (s)	-8	-4	0	+4	+8

The layer 3B inhibitory interneurons receive excitatory input from layer 4 and inhibitory input from all other inhibitory interneurons that correspond to the same position and disparity. Their cell membrane potentials, $q_{ijkd}^{L/R,+/-}$, are determined by the following equations:

$$\frac{d}{dt}q_{ijkd}^{L,+/-} = -\gamma_2 q_{ijkd}^{L,+/-} + \left[s_{(i+s)jk}^{L,+/-} - \theta\right]^+ - \beta\left(\left[q_{ijkd}^{R,+/-}\right]^+ + \left[q_{ijkd}^{R,-/+}\right]^+ + \left[q_{ijkd}^{L,-/+}\right]^+\right), \qquad \text{(A11)}$$

and

$$\frac{d}{dt}q_{ijkd}^{R,+/-} = -\gamma_2 q_{ijkd}^{R,+/-} + \left[s_{(i-s)jk}^{R,+/-} - \theta\right]^+ - \beta\left(\left[q_{ijkd}^{L,+/-}\right]^+ + \left[q_{ijkd}^{L,-/+}\right]^+ + \left[q_{ijkd}^{R,-/+}\right]^+\right), \qquad \text{(A12)}$$

where γ_2, β and θ are constants (4.5, 4, 0.4) representing the decay rate of the membrane potential, the strength of the inhibition and the signal threshold.

The exact values of parameters α, β, γ_1 and γ_2 are not critical. Under mild constraints on these parameters, the interactions of the binocular cells and interneurons make the binocular cells act like the "obligate cells" of Poggio (1991), responding vigorously only when their left and right inputs are approximately equal in magnitude. Equation (A10) was solved at equilibrium, using the equations described in the Obligate Theorem of Appendix B to speed up the simulations.

V1 Layer 2/3 monocular and binocular complex cells. V1 layer 2/3 consists of both monocular and binocular complex cells, which pool the cell membrane potentials of monocular/binocular layer 3B simple cells of like orientation and both contrast polarities at each position. These complex pyramidal cells also emit long-range, collinear, coaxial connections within layer 2/3 whereby they excite each other, as well as short-range, disynaptic interneurons that inhibit target complex cells as well as nearby inhibitory interneurons. This balance of excitation and inhibition helps to implement a bipole property that controls boundary grouping (Grossberg and Mingolla, 1985a, 1985b; Grossberg and Raizada, 2000). The monocular and

binocular collinear bipole cells obey the same dynamics but have different inputs. The equation for the membrane potential, c_{ijkd}, of binocular collinear bipole cell in V1 layer 2/3 is:

$$\frac{d}{dt}c_{ijkd} = -\alpha c_{ijkd} + (\beta - c_{ijkd})\left(I^{c}_{ijkd}\left(\gamma_1 + \gamma_2\left[\sum_{v} H^{E_c}_{ijkdv} - H^{I_c}_{ijkd} \right]^{+} \right) + \gamma_3\left[c_{ijkd} - \beta_c\right]^{+} \right) \quad \text{(A13)}$$
$$- (1 + c_{ijkd})(C^{P}_{ijkd} + C^{S}_{ijkd}),$$

where α , β, γ are constants ($\alpha = 20, \beta = 7$, $\gamma_1 = 1$, $\gamma_2 = 1$, $\gamma_3 = 0.5$). In (A13), I^{c}_{ijkd} is the input from V1 layer 3B binocular simple cells and is given by:

$$I^{c}_{ijkd} = \mu\left(\left[b^{B,+}_{ijkd} - \theta\right]^{+} + \left[b^{B,-}_{ijkd} - \theta\right]^{+}\right), \quad \text{(A14)}$$

where μ and θ are constants (20, 0.1). Term H^{Ec}_{ijkdv} in (A13) describes excitatory input from the long-range connections in V1 layer 2/3 to a complex cell at position (*i,j*), orientation *k* and disparity *d* (Grossberg and Swaminathan, 2004). Index *v* means that the input comes from side *v* of the complex cell. The term $\sum_{v} H^{E_c}_{ijkdv}$ in (A13) sums inputs from both sides *v=1,2* of the complex cell. In particular:

$$H^{Ec}_{ijkdv} = \sum_{pq} W^{c}_{pqijkv}[c_{pqkd} - \zeta_c]^{+}, \quad \text{(A15)}$$

where ζ_c is a constant threshold (0), and W^{c}_{pqijkv} is the long-range connection weight at orientation *k* and side *v* from the bipole cell at position (*p,q*) to the bipole cell at position (*i,j*). The connection weights for the horizontal orientation (*k=2*) are defined as follows (*v=1* for left branch and *v=2* for right branch):

$$W^{c}_{pqij21} = \left[sign(i-p)\exp\left(-\left(\frac{(i-p)^2}{\sigma_p^2} + \frac{(j-q)^2}{\sigma_q^2} \right) \right) \right]^{+}, \quad \text{(A16)}$$

and

$$W^{c}_{pqij22} = \left[sign(p-i)\exp\left(-\left(\frac{(i-p)^2}{\sigma_p^2} + \frac{(j-q)^2}{\sigma_q^2} \right) \right) \right]^{+}, \quad \text{(A17)}$$

where $sign(x) = 1$ if $x > 0$, -1 if $x < 0$, and 0 otherwise. The parameters $\sigma_p = 8$, $\sigma_q = 0.3$, and the spatial connection range (diameter) is 3. The connection weights for the vertical orientation are obtained by rotation.

Term H^{Ic}_{ijkd} in (A13) is the inhibitory input from the inhibitory interneurons, defined by:

$$H^{Ic}_{ijkd} = \sum_{v} [s^{c}_{ijkdv}]^{+}, \quad \text{(A18)}$$

where s^{c}_{ijkdv} represents the activity of an inhibitory interneuron associated with the complex cell at position *(i,j)*, orientation *k,* disparity *d*, and side *v*. There is an inhibitory interneuron associated with each side of a complex cell. This inhibition from the inhibitory interneurons helps to

maintain the balance between excitation and inhibition to enforce the bipole property. Potential s^c_{ijkdv} is defined by

$$\frac{d}{dt} s^c_{ijkdv} = \delta_I (-s^c_{ijkdv} + H^{E_c}_{ijkdv} - \eta s^c_{ijkdv} [s^c_{ijkdu}]^+), \qquad \text{(A19)}$$

where *u, v* are the two branches of orientation *k*, δ_I is a large constant representing the rapid response of the inhibitory interneuron, and $\eta = 1$. Equations (A13), (A18), and (A19) say that inhibitory interneurons inhibit each other as well as their target complex cells. Solving both the equations for s^c_{ijkdv} and s^c_{ijkdu} together at steady state, one finds:

$$s^c_{ijkdv} = (-B_v + \sqrt{B_v{}^2 + 4\eta H^{Ec}_{ijkdv}}) / 2\eta, \qquad \text{(A20)}$$

and

$$s^c_{ijkdu} = (-B_u + \sqrt{B_u{}^2 + 4\eta H^{Ec}_{ijkdu}}) / 2\eta, \qquad \text{(A21)}$$

where $B_v = 1 + \eta(H^{Ec}_{ijkdu} - H^{Ec}_{ijkdv})$ and $B_u = 1 + \eta(H^{Ec}_{ijkdv} - H^{Ec}_{ijkdu})$. These steady-state values are used in the simulations.

A collinear bipole cell has two orientation branches (e.g., left and right for the horizontal orientation), either of which is associated with one inhibitory interneuron. Therefore, there are two inhibitory interneurons associated with each collinear bipole cell. In (A19), only the related two inhibitory cells inhibit each other. A collinear bipole cell will not fire when it receives excitatory input from only one side of its long-range connection, but it can fire when it receives excitatory inputs from both sides. In (A19), for example, when H^{Ec}_{ijkdu} equals zero, then s^c_{ijkdu} equals zero. As a result, s^c_{ijkdv} equals H^{Ec}_{ijkdv}. Therefore, the total excitatory input from the long-range connections equal the total inhibitory input from the inhibitory interneurons, and thereby together contribute nothing to fire a cell. But when both H^{Ec}_{ijkdu} and H^{Ec}_{ijkdv} are far from zero, the total excitatory inputs are much larger than the inhibitory inputs, which helps to fire a bipole cell. In particular, a collinear bipole cell which receives large excitatory inputs from both sides of its long-range connection will fire if it is in V2. In V1, this model assumes that the excitatory inputs from long-range connections of a bipole cell are modulatory, which help to fire the bipole cell only if it also receives excitatory input from layer 4.

Term $\gamma_3 [c_{ijkd} - \beta_c]^+$ in (A13) is self-excitatory feedback, with threshold $\beta_c = 0.03$. Term C^P_{ijkd} is the inhibitory input at the same position and disparity from other bipole cells with different orientations, and is defined by

$$C^P_{ijkd} = \gamma_4 \left(\sum_{r \neq k} [c_{ijrd} - \beta_c]^+ \right), \qquad \text{(A22)}$$

where $\gamma_4 = 5$. Term C^S_{ijkd} is the inhibitory input from spatial competition across position and orientation but within disparity. Competition across position helps to sharpen the spatial positions of boundaries. Competition across orientation helps to prevent abutting lines from generating illusory contours that can penetrate interior lines of a different orientation (Spatial impenetrability; Grossberg and Mingolla, 1987; Grossberg and Williamson, 2001). C^S_{ijkd} is defined by

$$C^S_{ijkd} = \gamma_5 \sum_{p\neq i, q\neq j, r} W_{pqijk}[c_{pqrd} - \beta_c]^+ , \qquad \text{(A23)}$$

where $\gamma_5 = 1$, and W_{pqijk} is an elliptic Gaussian kernel elongated at the orientation perpendicular to orientation *k*. The kernel W_{pqijk} for vertical orientation (*k=1*) is defined as follows

$$W_{pqij1} = \exp\left(-\left(\frac{(i-p)^2}{\sigma_p^2} + \frac{(j-q)^2}{\sigma_q^2}\right)\right), \qquad \text{(A24)}$$

where $\sigma_p > \sigma_q$ ($\sigma_p = 8$, $\sigma_q = 0.3$). The kernel for horizontal orientation is obtained by appropriate rotation.

The membrane potential, $c^{L/R}_{ijk}$, of a monocular complex cell obeys the same equation as binocular complex cell c_{ijkd}:

$$\frac{d}{dt}c^{L/R}_{ijk} = -\alpha c^{L/R}_{ijk} + (\beta - c^{L/R}_{ijk})\left(I^{c,L/R}_{ijk}\left(\gamma_1 + \gamma_2\left[\sum_v H^{E_c}_{ijkv} - H^{I_c}_{ijk}\right]^+\right) + \gamma_3\left[c^{L/R}_{ijk} - \beta_c\right]^+\right) \qquad \text{(A25)}$$
$$-(1 + c^{L/R}_{ijk})(C^P_{ijk} + C^S_{ijk}),$$

where $I^{c,L/R}_{ijk}$ is the input from V1 layer 3B monocular simple cells that is defined by:

$$I^{c,L/R}_{ijk} = \left[b^{L/R,+}_{ijk} - \theta\right]^+ + \left[b^{L/R,-}_{ijk} - \theta\right]^+ . \qquad \text{(A26)}$$

The parameters $\alpha = 20$, $\beta = 8$, and $\theta = 0.4$. The other parameter values are the same as in the binocular case.

V2 layer 4. In V2, virtually all cells are binocularly driven (Hubel and Livingstone, 1987) consistent with the model hypothesis that the left and right monocular inputs are combined in layer 4 of V2. Since the monocular inputs do not yet have a depth associated with them, they are added to all depth planes along their respective lines-of-sight. Initially, the membrane potential, v_{ijkd}, of a V2 layer 4 cell is thus:

$$v_{ijkd} = \alpha h(\left[c_{ijkd} - \theta\right]^+) + \beta\left(h(\left[c^L_{(i+s)jk} - \theta_m\right]^+) + h(\left[c^R_{(i-s)jk} - \theta_m\right]^+)\right), \qquad \text{(A27)}$$

where *s* is the positional shift between left and right eye input that is defined in Table 1; *h* is the signal function with *h(x)=1* if *x>0*, *0* otherwise; θ and θ_m are constants representing the signal thresholds (0.06, 0.3); and α and β are constants (2.6, 0.8) representing the strength of the binocular and monocular connections.

The V2 layer 4 cells also receive feedback signals from the left and right V2 monocular surfaces (to be defined later) operating from V2 thin stripes to pale stripes. These are the surface-to-boundary feedback signals that were discussed in constraint (7) of Section 1. These signals modulate the activities of V2 layer 4 boundary cells. At steady-state, v_{ijkd} is defined by:

$$v_{ijkd} = \left(\alpha h(\left[c_{ijkd} - \theta\right]^+) + \beta\left(h(\left[c^L_{(i+s)jk} - \theta_m\right]^+) + h(\left[c^R_{(i-s)jk} - \theta_m\right]^+)\right)\right)(1 + \qquad \text{(A28)}$$
$$\alpha_f f_{ijkd})\left(\delta + (1-\delta)h(f_{ijkd})\right),$$

where f_{ijkd} is the total surface-to-feedback signal, α_f is a constant (1.1) that scales the strength of surface-to-boundary feedback signals, and α, β, *h*, θ and θ_m are the same as in (A27). The parameter δ is a constant (0.2) that scales the activities of layer 4 cells. A choice of $\delta < 1$ implies

that the activities of layer 4 cells that do not receive surface-to-boundary feedback signals are suppressed to some degree. The surface-to-boundary feedback signal f_{ijkd} is given by

$$f_{ijkd} = [f^L_{ijkd} - \theta_f]^+ + [f^R_{ijkd} - \theta_f]^+, \quad \text{(A29)}$$

where f^L_{ijkd} and f^R_{ijkd} are the left and right eye surface-to-boundary feedback signals, respectively, and θ_f is a constant signal threshold (0.03).

V2 layer 2/3 complex cells. The V2 layer 2/3 cells receive input from V2 layer 4. The bipole cells in V2 layer 2/3 implement perceptual grouping by long-range horizontal connections. The model also proposes the disparity filter can be realized as part of the inhibitory interactions that control perceptual grouping. See constraint (5) of Section 1. The membrane potential, g_{ijkd}, of the bipole cell in V2 layer 2/3 at position (i,j) that codes orientation k and disparity d obeys the equation:

$$\frac{d}{dt} g_{ijkd} = -\alpha g_{ijkd} + (\beta - g_{ijkd}) \left(\gamma_1 I^g_{ijkd} + \gamma_2 \left[\sum_v H^{E_g}_{ijkdv} - H^{I_g}_{ijkd} \right]^+ \right) - (1 + g_{ijkd}) G^P_{ijkd}, \quad \text{(A30)}$$

where α , β, γ_1 and γ_2 are constants ($\alpha = 30,\ \beta = 10$, $\gamma_1 = 1.4, \gamma_2 = 1$).

Term I^g_{ijkd} in (A30) is the input signal from V2 layer 4 that is given by:

$$I^g_{ijkd} = [v_{ijkd}]^+ . \quad \text{(A31)}$$

The V2 layer 2/3 collinear bipole cells receive long-range input from other (almost) collinear and coaxial bipole cells at nearby positions with the same disparity preference. Term $H^{E_g}_{ijkdv}$ is the input from branch *v* of the bipole cell at position *(i,j)*, orientation *k* and disparity *d*:

$$H^{Eg}_{ijkdv} = \sum_{pq} W^g_{pqijkv} [g_{pqkd} - \zeta_g]^+ . \quad \text{(A32)}$$

The long-range connection weight W^g_{pqijkv} is the same as W^c_{pqijkv} in (A16) and (A17), but with a larger spatial range (diameter = 7) and $\sigma_p = 15$, $\sigma_q = 0.1$, and $\zeta_g = 0.03$. Term H^{Ig}_{ijkd} is the inhibitory input from the inhibitory interneurons, defined by:

$$H^{Ig}_{ijkd} = \sum_v [s^g_{ijkdv}]^+ , \quad \text{(A33)}$$

with the activity, s^g_{ijkdv}, of the inhibitory interneuron for branch *v* being defined by:

$$\frac{d}{dt} s^g_{ijkdv} = \delta_I (-s^g_{ijkdv} + H^{E_g}_{ijkdv} - \eta s^g_{ijkdv} [s^g_{ijkdu}]^+) , \quad \text{(A34)}$$

where *u, v* are the two branches of orientation *k*, and parameters δ_I and η are the same as in (A19).

Each V2 layer 2/3 bipole cell also receives inhibitory input from other bipole cells that share either of its monocular inputs (line-of-sight competition). Term G^P_{ijkd} in (A30) is the inhibition across disparities along the lines-of-sight:

$$G^P_{ijkd} = \gamma_3 \sum_{d' \neq d} M_{dd'} ([g_{(i+s'-s)jkd'} - \beta_g]^+ + [g_{(i+s-s')jkd'} - \beta_g]^+), \quad \text{(A35)}$$

where γ_3 is a constant (5) which scales the total inhibition that each cell receives; $M_{dd'}$ is the strength of line-of-sight inhibition from all other cells that share a monocular input between disparities *d* and *d'* (see Table 2); and $[g_{(i+s'-s)jkd'} - \beta_g]^+$ and $[g_{(i+s-s')jkd'} - \beta_g]^+$ are V2 layer 2/3 bipole cell inhibitory inputs along the left and right lines-of-sight with positional shifts *s* and *s'* (see Table 1) and inhibitory signal threshold β_g (0.03).

Table 2. The inhibition coefficients $M_{dd'}$ that define line-of-sight inhibition. Each neuron is inhibited by every other neuron that shares either of its inputs by an amount that depends on the disparities of the inhibited and inhibiting neurons (cf. Figure 3).

	V. Near	**Near**	**Zero**	**Far**	**V. Far**
V. Near	-	3	5	3	2
Near	0.4	-	2.5	2	0.4
Zero	0.3	1.5	-	1.5	0.3
Far	0.4	2	2.5	-	0.4
V. Far	2	3	5	3	-

The disparity filter (G^P) in (A30) is robust in that its behavior is stable across a range of parameter values. The key features of the disparity filter, as illustrated in Table 2, is that it is symmetrical about the fixation plane (i.e., the near and far disparity planes equally inhibit and are equally inhibited by the zero disparity plane) and that it favors the zero disparity plane in that this plane inhibits the near and far disparity planes more than they inhibit it.

V2 thin stripe monocular surface filling-in. V2 thin stripes receive lightness signals from LGN via V1 blobs and binocular boundary signals from V2 layer 2/3 bipole cells to generate monocular surface filling-in. The boundary signals received from V2 layer 2/3 create resistive barriers to the filling-in process. Following Grossberg and Todorović (1988), the filling-in is modeled by a diffusion equation as follows:

$$\varepsilon \frac{d}{dt} F_{ijd}^{L/R} = -\alpha F_{ijd}^{L/R} + \sum_{(p,q)\in N_{ij}} (F_{pqd}^{L/R} - F_{ijd}^{L/R}) \Phi_{pqijd} + x_{ijd}^{L/R}, \qquad \text{(A36)}$$

where rate parameter $\varepsilon << 1$, represents that the monocular surface filling-in process in the thin stripes is faster than the grouping process in V2 layer 2/3; α is the decay rate (1.0); $F_{ijd}^{L/R}$ is the membrane potential of the left/right stripe cell at position (*i,j*) and disparity *d*; and N_{ij} is the set of the nearest-neighbor locations of *(i,j)*:

$$N_{ij} = \{(i, j-1), (i-1, j), (i+1, j), (i, j+1)\}; \qquad \text{(A37)}$$

and $x_{ijd}^{L/R}$ is the lightness signal from left/right LGN, given by:

$$x_{ijd}^{L} = [x_{(i+s)j}^{L}]^+, \qquad \text{(A38)}$$

and

$$x_{ijd}^{R} = [x_{(i-s)j}^{R}]^+, \qquad \text{(A39)}$$

where *s* is the positional shift defined in Table 1. The diffusion-gating coefficients, Φ_{pqijd}, in (A36) represent the inhibition of the diffusion by the boundary signals from V2 layer 2/3. They are defined by:

$$\Phi_{pqijd} = \frac{\delta}{1+\rho\left(g_{(i-0.5)(j-0.5)d} + g_{(i-0.5)(j+0.5)d}\right)}, \quad \text{if p = } i\text{-1 and q = } j, \quad \text{(A40)}$$

$$\Phi_{pqijd} = \frac{\delta}{1+\rho\left(g_{(i+0.5)(j-0.5)d} + g_{(i+0.5)(j+0.5)d}\right)}, \quad \text{if p = } i\text{+1 and q = } j, \quad \text{(A41)}$$

$$\Phi_{pqijd} = \frac{\delta}{1+\rho\left(g_{(i-0.5)(j-0.5)d} + g_{(i+0.5)(j-0.5)d}\right)}, \quad \text{if p = } i \text{ and q = } j\text{-1}, \quad \text{(A42)}$$

$$\Phi_{pqijd} = \frac{\delta}{1+\rho\left(g_{(i-0.5)(j+0.5)d} + g_{(i+0.5)(j+0.5)d}\right)}, \quad \text{if p = } i \text{ and q = } j\text{+1}, \quad \text{(A43)}$$

where the spread scale parameter $\delta = 2000$, and the blocking scale parameter $\rho = 200$. The boundary terms in these equations sum over all orientations of bipole cell activations:

$$g_{ijd} = \gamma \sum_k [g_{ijkd} - \theta_g]^+, \quad \text{(A44)}$$

with $\gamma = 10$ and signal threshold $\theta_g = 0.03$. Thus any large boundary value at the nearest neighbor locations reduces the diffusion coefficient Φ_{pqijd} and thereby blocks filling-in.

As shown by equations (A40)-(A43), the boundary lattice is offset by [0.5, 0.5] relative to the lightness lattice, corresponding to the idea that these two processing streams are spatially displaced with respect to one another in the cortical map. Spurious lattice edge effects were avoided by using the wrap-round technique according to which the last element of a row/column is adjacent to the first element of the same row/column (Grossberg and Howe, 2003).

Solving (A36) at steady-state yields:

$$F_{ijd}^{L/R} = \frac{x_{ijd}^{L/R} + \sum_{(p,q)\in N_{ij}} F_{pqd}^{L/R} \Phi_{pqijd}}{\alpha + \sum_{(p,q)\in N_{ij}} \Phi_{pqijd}}. \quad \text{(A45)}$$

This steady-state equation is used in the simulations.

In each time step of boundary grouping, monocular surfaces are filled-in and then generate surface-to-boundary feedback signals.

Surface-to-boundary feedback signals. The V2 monocular surfaces in the thin stripes generate surface-to-boundary feedback signals to the V2 pale stripes to modulate the activities of corresponding V2 layer 4 cells; see equation (A28). The simulations show how the surface-to-boundary feedback signals play a key role in explaining some data.

Output signals from the filled-in activities in the V2 thin stripes are derived from oriented filters

$$f_{ijkd}^{L/R,+} = \sum_{p,q} \mathrm{K}_{pqk} \left[F_{i+p,j+q,d}^{L/R}\right]^+, \quad \text{(A46)}$$

$$f_{ijkd}^{L/R,-} = -f_{ijkd}^{L/R,+} = -\sum_{p,q} \mathrm{K}_{pqk} \left[F_{i+p,j+q,d}^{L/R}\right]^+, \quad \text{(A47)}$$

where the Gabor kernel K_{pqk} is defined in equation (A7).

The surface-to-boundary signals $f_{ijkd}^{L/R}$ are finally defined by

$$f_{ijkd}^{L/R} = [f_{ijkd}^{L/R,+}]^+ + [f_{ijkd}^{L/R,-}]^+. \quad \text{(A48)}$$

V4. V4 receives lightness signals from the LGN via V1 blobs and V2 thin stripes, and boundary signals from V2 layer 2/3 via V1 interblob regions and V2 pale stripes. It combines the monocular lightness signals from the two eyes that correspond to the same 3D location. Its binocular lightness input, z_{ijd}, is given by:

$$z_{ijd} = \left[x^{L}_{(i+s)j}\right]^{+} + \left[x^{R}_{(i-s)j}\right]^{+}, \qquad \text{(A49)}$$

where *i, j* are positional indices, *d* the disparity and *s* the positional shift defined in Table 1. The V4 cell membrane potentials, w_{ijd}, undergo binocular surface filling-in that represents the visible percept. This process is modeled by a diffusion equation similar to (A36) that is solved at steady-state:

$$w_{ijd} = \frac{z_{ijd} + \sum\limits_{(p,q)\in N_{ij}} w_{pqd}\Phi_{pqijd}}{\alpha + \sum\limits_{(p,q)\in N_{ij}} \Phi_{pqijd}}, \qquad \text{(A50)}$$

where parameter α=1; N_{ij} is defined in (A37); and Φ_{pqijd} is defined in (A40)-(A43) with the spread scale parameter $\delta = 1000$ and the blocking scale parameter $\rho = 400$.

Appendix B. Obligate Theorem

The enhanced model uses a different activation equation (see Equation (A10)) for layer 3B simple cells from Grossberg and Howe (2003) model. Here a self-normalized factor is added, which is closer to real cell activation. The following theorem shows that the obligate property holds at the binocular simple cells in layer 3B.

Obligate Theorem. Consider the system:

$$\frac{db_{ijkd}^{B,+}}{dt} = -\gamma_1 b_{ijkd}^{B,+} + (1 - b_{ijkd}^{B,+})\left(S_{(i+s)jk}^{L,+} + S_{(i-s)jk}^{R,+}\right) - \alpha\left(\left[q_{ijkd}^{L,+}\right]^+ + \left[q_{ijkd}^{L,-}\right]^+ + \left[q_{ijkd}^{R,+}\right]^+ + \left[q_{ijkd}^{R,-}\right]^+\right), \tag{B1}$$

$$\frac{dq_{ijkd}^{L,+}}{dt} = -\gamma_2 q_{ijkd}^{L,+} + S_{(i+s)jk}^{L,+} - \beta\left(\left[q_{ijkd}^{R,+}\right]^+ + \left[q_{ijkd}^{R,-}\right]^+ + \left[q_{ijkd}^{L,-}\right]^+\right), \tag{B2}$$

$$\frac{dq_{ijkd}^{R,+}}{dt} = -\gamma_2 q_{ijkd}^{R,+} + S_{(i-s)jk}^{R,+} - \beta\left(\left[q_{ijkd}^{L,+}\right]^+ + \left[q_{ijkd}^{L,-}\right]^+ + \left[q_{ijkd}^{R,-}\right]^+\right), \tag{B3}$$

$$\frac{dq_{ijkd}^{L,-}}{dt} = -\gamma_2 q_{ijkd}^{L,-} + S_{(i+s)jk}^{L,-} - \beta\left(\left[q_{ijkd}^{R,-}\right]^+ + \left[q_{ijkd}^{R,+}\right]^+ + \left[q_{ijkd}^{L,+}\right]^+\right), \tag{B4}$$

and

$$\frac{dq_{ijkd}^{R,-}}{dt} = -\gamma_2 q_{ijkd}^{R,-} + S_{(i-s)jk}^{R,-} - \beta\left(\left[q_{ijkd}^{L,-}\right]^+ + \left[q_{ijkd}^{L,+}\right]^+ + \left[q_{ijkd}^{R,+}\right]^+\right), \tag{B5}$$

where

$$S_{(i+s)jk}^{L,+/-} = \left[s_{(i+s)jk}^{L,+/-} - \theta\right]^+, \tag{B6}$$

and

$$S_{(i-s)jk}^{R,+/-} = \left[s_{(i-s)jk}^{R,+/-} - \theta\right]^+, \tag{B7}$$

with $s_{(i+s)jk}^{L,+/-}$ and $s_{(i-s)jk}^{R,+/-}$ being monocular simple cell activities that are defined by (A6) and (A8), $\theta \geq 0$, $\gamma_1 > 0$ and

$$0 < \beta < \gamma_2 < \alpha < \gamma_2 + \beta \tag{B8}$$

Under these conditions, the system converges exponentially to the unique equilibrium specified by (1), (2), (3) and (4) provided that the inputs are constant. Let $\Gamma = \gamma_1 + S_{(i+s)jk}^{L,+} + S_{(i-s)jk}^{R,+}$.

(1) if $0 < S_{(i+s)jk}^{L,+}, S_{(i-s)jk}^{R,+}$; and $\frac{\beta}{\gamma_2} \leq \frac{S_{(i+s)jk}^{L,+}}{S_{(i-s)jk}^{R,+}} \leq \frac{\gamma_2}{\beta}$,

then at equilibrium

$$b_{ijkd}^{B,+} = \frac{1}{\Gamma}\left(1 - \frac{\alpha}{\gamma_2 + \beta}\right)\left(S_{(i+s)jk}^{L,+} + S_{(i-s)jk}^{R,+}\right); \tag{B9}$$

(2) if $0 < S_{(i+s)j}^{V,L,+}, S_{(i-s)j}^{V,R,+}$; and $\frac{S_{(i+s)jk}^{L,+}}{S_{(i-s)jk}^{R,+}} > \frac{\gamma_2}{\beta_2}$,

then at equilibrium

$$b_{ijkd}^{B,+} = \frac{1}{\Gamma}\left(S_{(i-s)jk}^{R,+} + \left(1 - \frac{\alpha}{\gamma_2}\right) S_{(i+s)jk}^{L,+} \right); \quad \text{(B10)}$$

(3) if $0 < S_{(i+s)j}^{V,L,+}, S_{(i-s)j}^{V,R,+}$; and $\frac{S_{(i+s)jk}^{L,+}}{S_{(i-s)jk}^{R,+}} < \frac{\beta}{\gamma_2}$,

then at equilibrium

$$b_{ijkd}^{B,+} = \frac{1}{\Gamma}\left(S_{(i+s)jk}^{L,+} + \left(1 - \frac{\alpha}{\gamma_2}\right) S_{(i-s)jk}^{R,+} \right); \quad \text{(B11)}$$

(4) for all other values of $S_{(i+s)jk}^{L,+}, S_{(i-s)jk}^{R,+}$,

at equilibrium $b_{ijkd}^{B,+} \le 0$. (B12)

The proof of the theorem is similar to the one described in Grossberg and Howe (2003) and thus omitted.

References

Abbott, L.F., Varela, K. Sen, K., and Nelson, S.B. (1997). Synaptic depression and cortical gain Control. *Science*, *275*, 220-223.

Bakin, J. S., Nakayama, K. and Gilbert, C. D. (2000). Visual responses in monkey area V1 and V2 to three-dimensional surface configurations. *The Journal of Neuroscience*, *20,* 8188-8198.

Bringuier, V., Chavane, F., Glaeser, L. and Frégnac, Y. (1999). Horizontal propagation of visual activity in the synaptic integration field of area 17 neurons. *Science, 283*, 695-699.

Cao, Y. and Grossberg, S. (2004). A laminar cortical model of stereopsis and 3D surface perception: da Vinci stereopsis and closure. *Proceedings of Eighth International Conference on Cognitive and Neural Systems, p. 68.*

Cohen, M.A. and Grossberg, S. (1984). Neural dynamics of brightness perception: Features, boundaries, diffusion, and resonance. *Perception and Psychophysics*, *36*, 428-456.

Crook, J.M., Engelmann, R. and Lowel, S. (2002). Caba-inactivation attenuates collinear facilitation in cat primary visual cortex. *Experimental Brain Research*, *143*, 295-302.

Fang, L. and Grossberg, S. (2004). How are the surface lightnesses of complex stereograms assigned to the correct depths? *Fourth annual meeting of Vision Sciences Society, F13.*

Francis, G., Grossberg, S., Mingolla, E. (1994). Cortical dynamics of feature binding and reset: Control of visual persistence. *Vision Research*, *34*, 1089-1104.

Gillam, B., Blackburn, S. and Cook, M. (1995). Panum's limiting case: double fusion, convergence error, or 'da Vinci stereopsis'. *Perception*, *24,* 333-346.

Gillam, B., Blackburn, S. and Nakayama, K. (1999). Stereopsis based on monocular gaps: Metrical encoding of depth and slant without matching contours. *Vision Research*, *39,* 493-502.

Grimson, W. E. (1981). A computer implementation of a theory of human stereo vision. *Philosophical Transactions of the Royal Society (B), 292,* 217-253.

Grossberg, S. (1968). Some physiological and biochemical consequences of psychological postulates. *Proceedings of the National Academy of Sciences*, *60,* 758-765.

Grossberg, S. (1969). On the production and release of chemical transmitters and related topics in cellular control. *Journal of Theoretical Biology*, *22,* 325-364.

Grossberg, S. (1973). Contour enhancement, short-term memory and constancies in reverberating neural networks. *Studies in Applied Mathematics*, *52*, 217-257.

Grossberg, S. (1976). Adaptive pattern classification and universal recoding, II: Feedback, expectation, olfaction, and illusions. *Biological Cybernetics*, *23,* 187-202.

Grossberg, S. (1980). How does a brain build a cognitive code? *Psychological Review*, *87,* 1-51.

Grossberg, S. (1984). Outline of a theory of brightness, color, and form perception. In E. Degreef and J. van Buggenhaut (Eds.), *Trends in mathematical psychology* (pp. 5559-5586). Amsterdam: North-Holland.

Grossberg, S. (1987). Cortical dynamics of three-dimensional form, color, and brightness perception: II. Binocular theory. *Perception and Psychophysics*, *41,* 117-158.

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

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

Grossberg, S. (1999a). How does the cerebral cortex work? Learning, attention and grouping by the laminar circuits of visual cortex, *Spatial Vision, 12*, 163-186.

Grossberg, S. (1999b). The link between brain learning, attention, and consciousness. *Consciousness and Cognition, 8,* 1-44.

Grossberg, S. (2000). The complementary brain: Unifying brain dynamics and modularity. *Trends in Cognitive Sciences*, *4*, 233-246.

Grossberg, S. and Cao, Y. (2004). A laminar cortical model of stereopsis and 3D surface perception: closure and da Vinci stereopsis. *Fourth annual meeting of the Vision Sciences Society, F30.*

Grossberg, S. and Hong, S. (2004). A neural model of surface perception: Lightness, anchoring, and filling-in. *Spatial Vision, in press.*

Grossberg, S. and Howe, P.D.L. (2003). A laminar cortical model of stereopsis and three-dimensional surface perception. *Vision Research, 43*, 801-829.

Grossberg, S. and Kelly, F. (1999). Neural dynamics of binocular brightness perception. *Vision Research, 39,* 3796-3816.

Grossberg, S. and McLoughlin, N.P. (1997). Cortical dynamics of 3-D surface perception: Binocular and half-occluded scenic images. *Neural Networks*, *10,* 1583-1605.

Grossberg, S. and Mingolla, E. (1985a). Neural dynamics of perceptual grouping: textures, boundaries, and emergent segmentations. *Perception and Psychophysics, 38,* 141-147.

Grossberg, S. and Mingolla, E. (1985b). Neural dynamics of form perception: Boundary completion, illusory figures, and neon color spreading. *Psychological Review*, *92*, 173-211.

Grossberg, S. and Mingolla, E. (1987). Neural dynamics of surface perception: Boundary webs, illuminants, and shape-from-shading. *Computer Vision, Graphics and Image processing*, *37*, 116-165.

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

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,* 1413-1432.

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,* 852-863.

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, 1147-1187.

Grossberg, S. and Todorović D. (1988). Neural dynamics of 1-D and 2-D brightness perception: A unified model of classical and recent phenomena. *Perception and Psychophysics, 43,* 241-277.

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 groupings and learning. *Cerebral Cortex, 11,* 37-58.

Grossberg, S. and Yazdanbakhsh, A. (2003a). Laminar cortical dynamics of 3-D surface stratification, transparency, and neon spreading [Abstract]. *Journal of Vision*, *3*(9), 247a.

Grossberg S. and Yazdanbakhsh A. (2003b). Laminar cortical mechanisms of 3D Surface Processing. *Society for Neuroscience 33rd annual meeting*, 339.5.

Grossberg S. and Yazdanbakhsh A. (2005). Laminar Cortical Dynamics of 3D Surface Perception: Stratification, Transparency, and Neon Color Spreading. *Vision Research, in press.*

Hodgkin, A. L. (1964). *The conduction of the nervous impulse*. Springfield, IL: Charles C. Thomas.

Howard, I. P. and Rogers, B. J. (1995). *Binocular Vision and Stereopsis.* New York: Oxford University Press.

Howe, P. D. L. and Watanabe, T. (2003). Measuring the depth induced by an opposite-luminance (but not anticorrelated) stereogram. *Perception, 32,* 415-421.

Hubel, D. H. and Livingstone, M. S. (1987). Segregation of form, color, and stereopsis in primate area 18. *The Journal of Neuroscience*, *7,* 3378-3415.

Hubel, D. H. and Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology*, *195,* 215-243.

Ito, M. and Gilbert, C.D. (1999). Attention modulates contextual influences in the primary visual cortex of alert monkeys. *Neuron, 22*, 593-604.

Julesz, B. (1971). *Foundations of Cyclopean Perception.* Chicago: The University of Chicago Press.

Kapadia, M.K., Ito, M., Gilbert, C.D. and Westheimer, G. (1995). Improvement in visual sensitivity by changes in local context: Parallel studies in human observers and in V1 of alert monkeys. *Neuron, 15*, 843-856.

Kelly, F. J. and Grossberg, S. (2000). Neural dynamics of 3-D surface perception: Figure-ground separation and lightness perception. *Perception and Psychophysics, 62*, 1596-1619.

Lamme, V. A. F., Rodriguez-Rodriguez, V. and Spekreijse, H. (1999). Separate processing dynamics for texture elements, boundaries and surfaces in primary visual cortex of the Macaque monkey. *Cerebral Cortex*, *9*(4), 406-413.

Marr, D. and Poggio, T. (1976). Cooperative computation of stereo disparity. *Science, 194,* 283-287.

McKee, S. P., Bravo, M. J., Smallman, H. S. and Legge, G. E. (1995). The 'uniqueness constraint' and binocular masking. *Perception*, *24,* 49-65.

McKee, S. P., Bravo, M. J., Taylor, D. G. and Legge, G. E. (1994). Stereo matching precedes dichoptic masking. *Vision Research*, *34,* 1047-1060.

McLoughlin, N.P. and Grossberg, S. (1998). Cortical computation of stereo disparity. *Vision Research*, *38*, 91-99.

Nakayama, K. and Shimojo, S. (1990). da Vinci stereopsis: depth and subjective occluding contours from unpaired image points. *Vision Research*, *30,* 1811-1825.

Ohzawa, I., DeAngelis, G. C. and Freeman, R. D. (1990). Stereoscopic depth discrimination in the visual cortex: Neurons ideally suited as disparity detectors. *Science*, *249,* 1037-1041.

Pinna, B. and Grossberg, S. (2005). The watercolor illusion and neon color spreading: A unified analysis of new cases and neural mechanisms. Boston University Technical Report CAS/CNS TR-2005-003. Submitted for publication.

Pinna, B., Brelstaff, G., and Spillmann, L. (2001). Surface color from boundaries: A new 'watercolor' illusion. *Vision Research, 41*, 2669-2676.

Panum, P. L. (1858). *Physiologische Untersuchungen ueber das Sehen mit zwei Augen. Kiel:* Schwerssche Buchhandlung), translated by C Hubscher 1940 (Hanover, NH: Dartmouth Eye Institute).

Paradiso, M. A. and Nakayama, K. (1991). Brightness perception and filling-in. *Vision Research*, *31*, 1221-1236.

Pessoa, L. and Neumann, H. (1998). Why does the brain fill-in? *Trends in Cognitive Sciences*, *2*(11), 422-424.

Pessoa, L., Thompson, E. and Noë, A. (1998). Finding out about filling-in: a guide to perceptual completion for visual science and the philosophy of perception. *Behavioral and Brain Sciences*, *21*(6), 723-802.

Poggio, G. F. (1991). Physiological basis of stereoscopic vision. In: *Vision and Visual Dysfunction. Binocular vision* (pp. 224-238). Boston, MA: CRC.

Polat, U., Mizobe, K., Pettet, M.W., Kasamatsu, T. and Norcia, A.M. (1998). Collinear stimuli regulate visual responses depending on cell's contrast threshold. *Nature, 391*, 580-584.

Raizada, R. and Grossberg, S. (2001). Context-sensitive bindings by the laminar circuits of V1 and V2: A unified model of perceptual grouping, attention, and orientation contrast. *Visual Cognition, 8*, 431-466.

Ramachandran, V.S. and Nelson, J.I. (1976). Global grouping overrides point-to-point disparities. *Perception, 5*, 125-128.

Roelfsema, P.R., Lamme, V. A.F. and Spekreijse, H. (1998). Object-based attention in the primary visual cortex of the Macaque monkey. *Nature, 395*, 376-381.

Roelfsema, P.R. and Spekreijse, H. (1999). Correlates of gradual spread of attention over a traced curve in Macaque area V1. In *Society for Neuroscience Abstr., 7.2*.

Rossi, A. F., Rittenhouse, C. D. and Paradiso, M. A. (1996). The representation of brightness in primary visual cortex. *Science*, *273*, 1104-1107.

Schor, C. and Heckmann, T. (1989). Interocular differences in contrast and spatial frequency: effects on stereopsis and fusion, *Vision Research, 29,* 837-847.

Smallman, H. S. and McKee, S. P. (1995). A contrast ratio constraint on stereo matching. *Proceedings of the Royal Society of London B*, *260,* 265-271.

Tse, P. U. (2005). Voluntary attention modulates the brightness of overlapping transparent surfaces. *Vision Research, in press*.

Tausch, R. (1953). Die beidaugige Raumwahmehmung-ein Prozess auf Grund der Korrespondenz und Disparation von Gestalten anstelle der Korrespondenz oder Disparation einzelner Netzhautelemente. *Zeitschrift fur Experimentelle und Angewandte Psychologie, 1*, 394-421.

von der Heydt, R. and Peterhans, E. (1989). Mechanisms of contour perception in monkey visual cortex. I. lines of pattern discontinuity. *Journal of Neuroscience*, *9*, 1731-1748.

von der Heydt, R., Peterhans, E. and Baumgartner, G. (1984). Illusory contours and cortical neuron responses. *Science, 224*, 1260-1262.

Wilde, K. (1950). Der Punktreiheneffekt und die Rolle der binocularen Querdisparation beim Tienfenshen. *Psychologische Forschung, 23*, 223-262.

Yazdanbakhsh, A. and Grossberg, S. (2005). Laminar cortical dynamics of binocular rivalry. *Proceedings of the 2005 annual meeting of the Vision Sciences Society.*

Yazdanbakhsh A, Watanabe T. (2004). Asymmetry between horizontal and vertical illusory lines in determining the depth of their embedded surface, *Vision Research, 44*, 2621-2627.