Stereopsis and 3D Surface Perception by Spiking Neurons in Laminar Cortical Circuits: A Method for Converting Neural Rate Models into Spiking Models

Yongqiang Cao and Stephen Grossberg

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

Running title: Spiking laminar cortical model of depth perception

Neural Networks, in press

Submitted: December 28, 2010
Revised: October 16, 2011

All correspondence should be addressed to
Professor Stephen Grossberg
Center for Adaptive Systems
Boston University
677 Beacon Street
Boston, MA 02215
Phone: 617-353-7858
Fax: 617-353-7755
Email:steve@bu.edu

Acknowledgements
This work was supported in part by CELEST, a National Science Foundation Science of Learning Center (SBE-0354379), and by the SyNAPSE program of DARPA (HR0011-09-C-0001).

Abstract

A laminar cortical model of stereopsis and 3D surface perception is developed and simulated. The model shows how spiking neurons that interact in hierarchically organized laminar circuits of the visual cortex can generate analog properties of 3D visual percepts. 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 suggests how surface-to-boundary feedback from V2 thin stripes to pale stripes helps to explain how computationally complementary boundary and surface formation properties lead to a single consistent percept, eliminate redundant 3D boundaries, and trigger figure-ground perception. The model also shows how false binocular boundary matches may be eliminated by Gestalt grouping properties. 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 3D sLAMINART model simulates 3D surface percepts that are consciously seen in 18 psychophysical experiments. These percepts include 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, da Vinci stereopsis, and perceptual closure. The model hereby illustrates a general method of unlumping rate-based models that use the membrane equations of neurophysiology into models that use spiking neurons, and which may be embodied in VLSI chips that use spiking neurons to minimize heat production.

Running Title: Spiking laminar cortical model of depth perception

Keywords: visual cortex; stereopsis; depth perception; binocular vision; perceptual grouping; surface perception; lightness perception; spiking neurons; laminar cortical circuits; monocular-binocular interactions; V1; V2; V4; LAMINART model; spiking neurons; integrate-and-fire neuron; VLSI

1. Introduction: From rate-based models to spiking models of brain and behavior

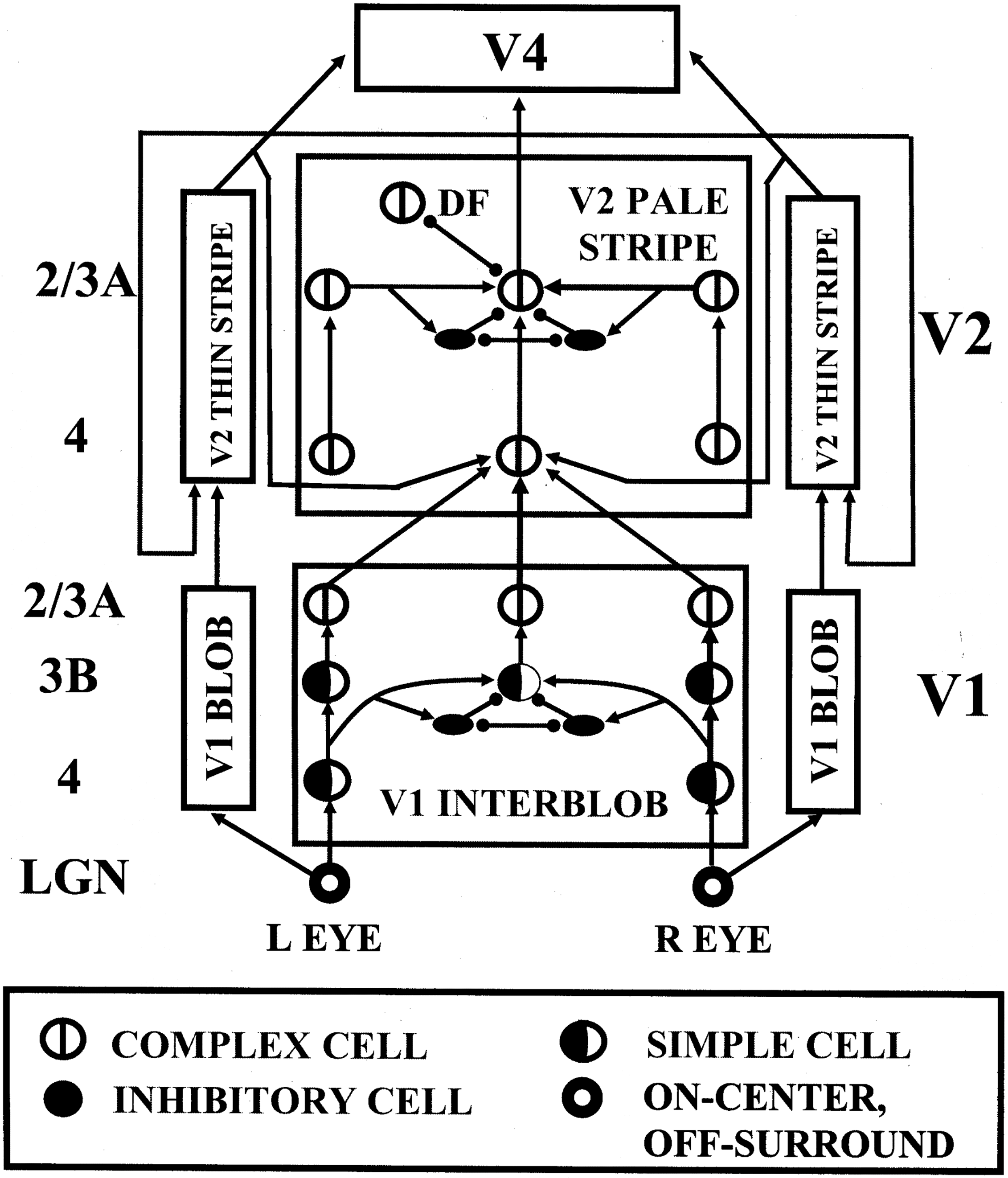
1.1. Explaining 3D percepts with laminar cortical models of spiking neurons. This article unifies three major themes in visual perception. First is the theme of how the brain perceives the world in depth using representations of 3D boundaries and surfaces. Second is the theme of how and why visual cortex and other neocortical structures are hierarchically organized into layered circuits. Third is the theme that neurons in the brain often communicate using discrete spikes. This article proposes how the brain uses discrete spikes in hierarchically organized laminar cortical circuits to generate 3D boundaries and surfaces with analog properties that match visual percepts. The current model hereby contributes to the goal of processing natural scenes using laminar circuits of spiking neurons.

1.2. Converting rate models into spiking models for implementation in VLSI chips. A theme of related interest is that a new generation of ultra-compact VLSI chips is being developed to emulate properties of brain dynamics; cf., the DARPA SyNAPSE program. These chips often use spiking neurons to communicate across networks of neurons in order to use less power and thereby minimize the production of heat. Nodes in these chips are also often organized in layers to facilitate inter-node communication.

Cao and Grossberg (2005) developed a laminar cortical model of stereopsis and 3D surface perception using rate-based neurons to quantitatively simulate 18 challenging psychophysical experiments, a feat still not matched by competing models, and did so with a single set of model parameters. Every cell type in the 3D LAMINART model is supported by anatomical and neurophysiological data. The current article shows how the same challenging set of psychophysical data can be simulated with a laminar cortical model that is composed of spiking neurons. We call this enhanced model the *3D spiking LAMINART* (sLAMINART) model (Figure 1). Our new results illustrate a general method for converting neural rate models that use the membrane, or shunting, equations of neurophysiology (Grossberg, 1973, 1980; Hodgkin, 1964) into spiking models that preserve the functional properties of these rate models, including their capacity to exhibit analog properties. This method opens the way to translating almost a half-century of neural models that use such shunting models (e.g., http://cns.bu.edu/~steve) into spiking models that can be embodied in VLSI chips for multiple applications.

Rate models are useful for several reasons: First, they are conceptually easier to discover, develop, and test than spiking models. Second, simulating them makes far less computational demands on computer resources. This fact alone made the use of rate models imperative to explain non-trivial properties of behavior when computer power was historically more limited. Third, and most importantly, rate models are often capable of quantitatively simulating large behavioral and brain data bases, and qualitatively explaining even larger data bases. Such a pragmatic test of a model suggests that spiking dynamics are not rate-limiting in the explanation of many data, if only because the brain often averages over spikes in space and time. This article hereby illustrates a research strategy in which a rate-based model of an area of biological intelligence is first developed and validated by explaining and simulating a large interdisciplinary data base. Then, once the design principles and mechanisms of the rate-based model are well understood, one can follow the type of procedure demonstrated in this article to convert it into a spiking model for possible implementation in spiking VLSI chips.

a)


V4
DF
V2 PALE STRIPE
V2 THIN STRIPE
V2 THIN STRIPE
2/3A
4
V2
2/3A
3B
4
LGN
V1 BLOB
V1 BLOB
V1 INTERBLOB
V1
L EYE
R EYE
COMPLEX CELL
SIMPLE CELL
INHIBITORY CELL
ON-CENTER, OFF-SURROUND

b)

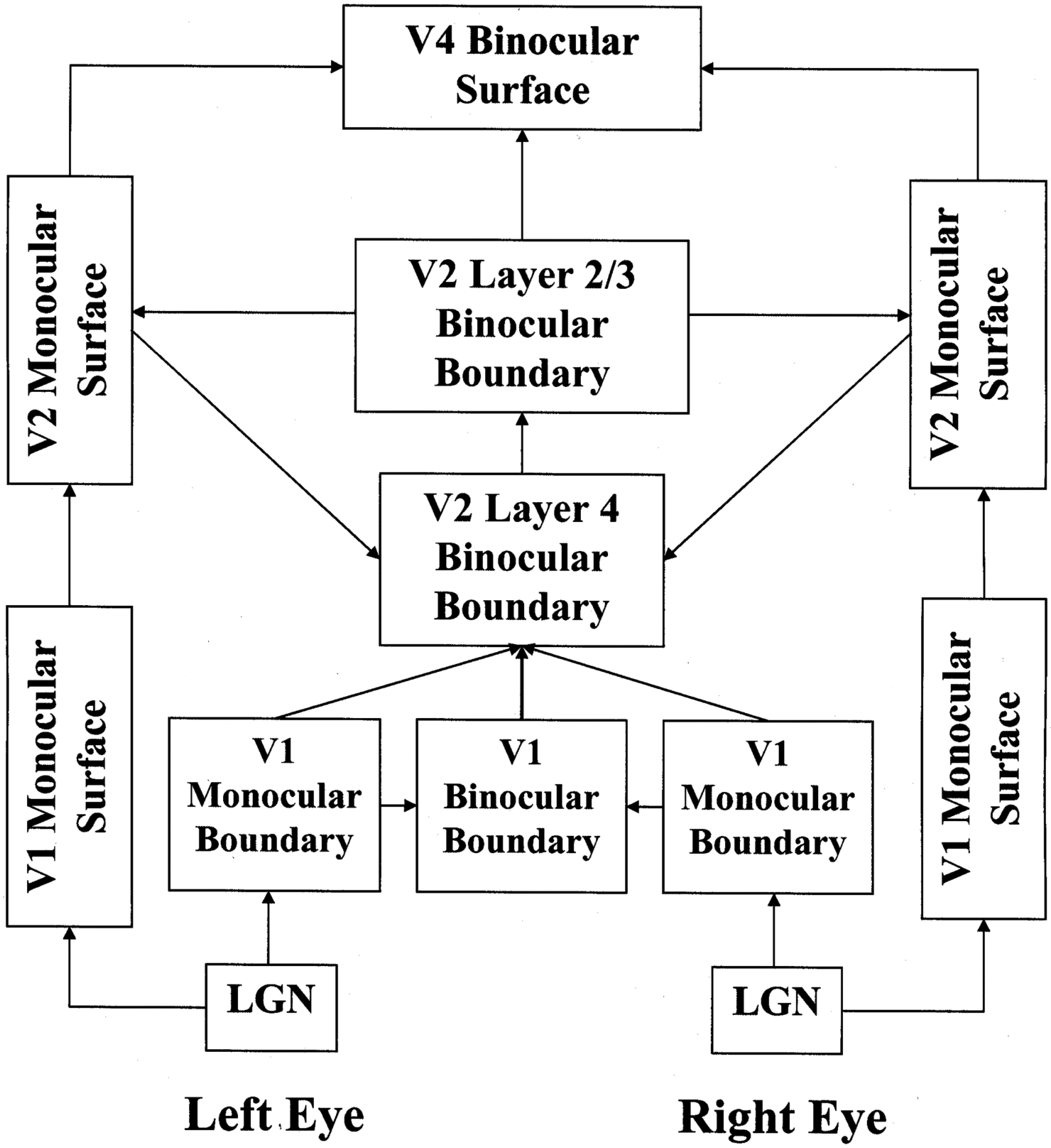


Figure 1. (a) The 3D LAMINART model circuit diagram. The model consists of a (V1 Interblob) - (V2 Pale Stripe)-V4 boundary stream which computes 3D perceptual groupings, and a (V1 Blob) - (V2 Thin Stripe)-V4 surface stream which computes 3D surface representations of lightness, color, and depth. The two processing streams interact to overcome their complementary deficiencies (Grossberg, 2000) and create consistent 3D boundary and surface percepts. (b) A block diagram of the enhanced 3D LAMINART model. [Reproduced with permission from Cao and Grossberg (2005).]

The above comments to not deny that there are cases where spiking models exhibit properties that do go beyond the predictive scope of rate models. Grossberg and Versace (2008), for example, developed a model in which spiking dynamics, local field potentials, synchronous oscillations, and spike-timing-dependent plasticity were implemented in a hierarchy of laminar cortical circuits interacting with specific and nonspecific thalamic nuclei for the purpose of learning recognition categories. This Synchronous Matching Adaptive Resonance Theory, or SMART, model made predictions about how fast gamma oscillations may be triggered when a top-down expectation matches bottom-up data, how slower beta oscillations may be triggered when a top-down expectation does not match bottom-up data, and how modulation of the selectivity of learned recognition categories by a vigilance process may be controlled by acetylcholine release when the nucleus basalis of Meynert is activated by a mismatch. Knowing when to use rate-based or spike-based models, and how to convert between them, is needed to realize a mature computational neuroscience and neuromorphic technology.

1.3. Seven organizational principles for explaining how the brain sees in depth. The 3D sLAMINART model, likes its rate-based precursor, achieves its explanatory goal by embodying seven basic constraints in its laminar cortical circuitry (see Figure 1):

(1) *Reconciles contrast-specific binocular fusion with contrast-invariant boundary perception*. 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; Howe and Watanabe, 2003). Binocular fusion thus obeys the *same-sign hypothesis* (Figure 2a). However, fused boundaries must also be able to form around objects whose contrast polarity with respect to the background can reverse along their perimeters (Figure 2b; Grossberg, 1994). Thus, binocular boundaries also 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 emerge through interactions between cells in layers 4, 3B, and 2/3A of cortical area V1 interblobs (Figure 1).

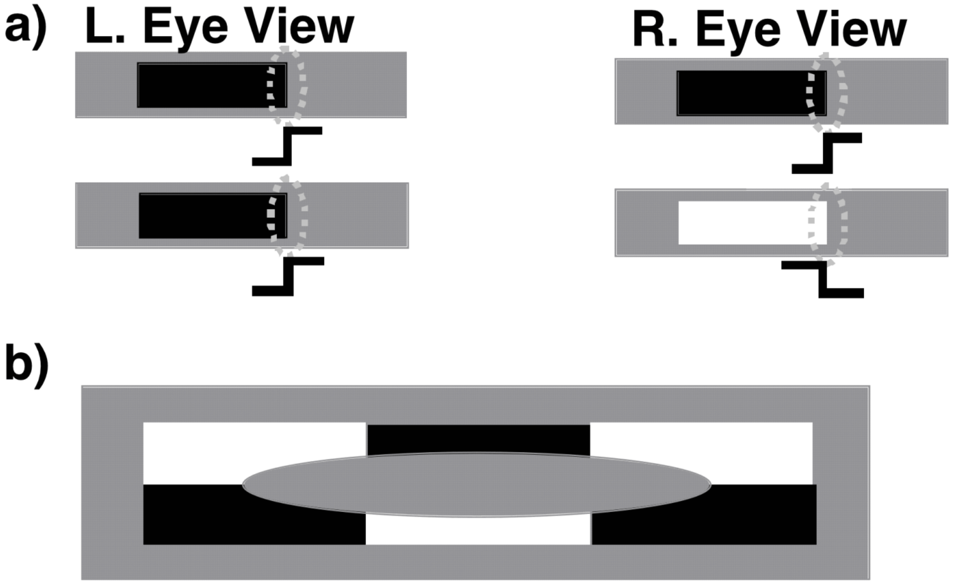


Figure 2. (a) The same-sign hypothesis: Only edges that have the same contrast polarity can be stereoscopically fused to produce a percept of depth; (b) As it is traversed, the boundary of the ellipse changes its contrast polarity relative to the background, thereby illustrating the need for object boundaries to be represented in a contrast-invariant manner. [Reproduced with permission from Cao and Grossberg (2005).]

(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 (Figure 1a) 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.

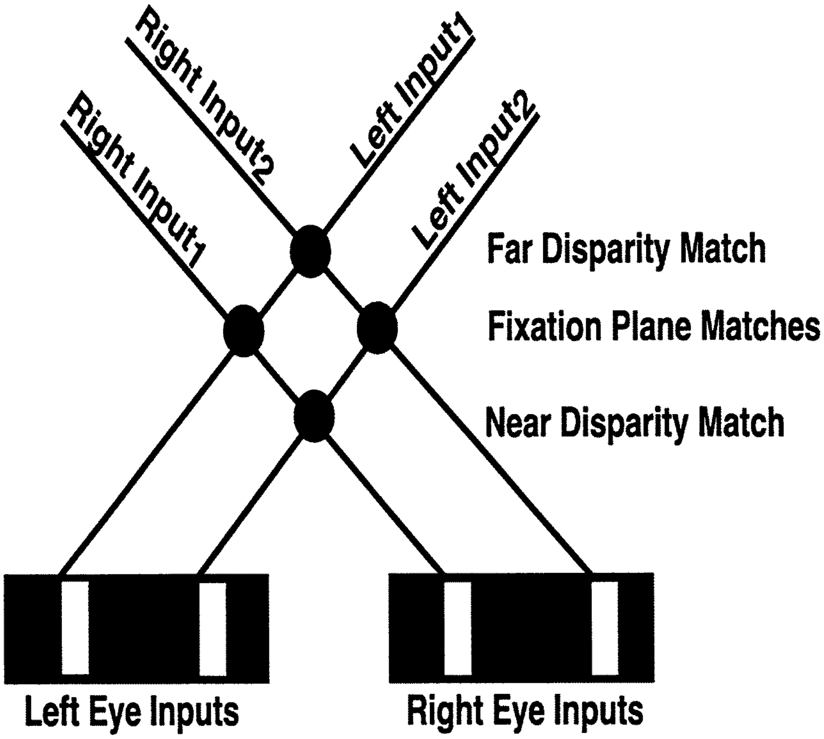


Figure 3. The V2 disparity filter. The V1 binocular boundary network matches an edge in one retinal image with every other edge in the other retinal image whose relative disparity is not too great, that has the same contrast polarity, and whose magnitude of contrast is not too different. In response 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 allelotropic shifts (see Table 1): an edge in the left retinal image is shifted to the right for matches increasingly farther away, whereas an edge in the right retinal image is shifted in the opposite direction. [Reproduced with permission from Cao and Grossberg (2005).]

(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 (Figure 3). This Correspondence 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 sLAMINART model does not enforce unique matches. It encourages them by using a *disparity filter* (Grossberg and McLoughlin, 1997; McLoughlin and Grossberg, 1998). The disparity filter uses inhibition between active cells that represent different depths in order to encourage unique matches. These inhibitory interactions are proposed to occur among the inhibitory connections that carry out the perceptual grouping process within layer 2/3 (see item (5) below).

(4) *Combines monocular and binocular information to form depth percepts*. Panum's limiting case has homologs during perception of the real world. Indeed, 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 by a scenic object 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, does the brain determine to what 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, and the same hypothesis was shown by Grossberg and Howe (2003) to help explain many data about 3D surface perception. This hypothesis proposes 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 using its inhibitory interactions to eliminate 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 excitation along long-range horizontal connections 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 hereby 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 inducing stimuli are present that are collinearly aligned on both sides of a target cell, 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 inhibit the target cell, as well as each other via recurrent inhibitory connections (Grossberg and Raizada, 2000). This recurrent inhibition normalizes the total amount of inhibition emanating from the interneuron pool. The summating excitation and normalizing inhibition create a case of "two-against-one" so that the target cell may be 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. In the 3D sLAMINART model, some of these inhibitory interneurons also realize the disparity filter as part of the grouping process. In other words, we propose that *the selection of a correct 3D grouping includes the suppression of false binocular matches*. This hypothesis links a solution of the Correspondence Problem to the Gestalt grouping problem.

Several lines of psychophysical data may be parsimoniously explained by this hypothesis. First, it enables the model to explain and simulate many psychophysical data about stereopsis and 3D surface perception. Second, it clarifies the role of perceptual grouping in depth perception: 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, it can also happen that 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 data about how the 3D grouping process can suppress false matches that do not correspond to the correct image matches from the outside world.

The role of perceptual groupings in determining conscious 3D percepts is also illustrated by a variety of other data about 3D vision, including data about bistable perception and binocular rivalry (Grossberg and Swaminathan, 2004; Grossberg, Yazdanbakhsh, Cao, and Swaminathan, 2008) and about perceptual transparency and stratification (Grossberg and Yazdanbakhsh, 2005).

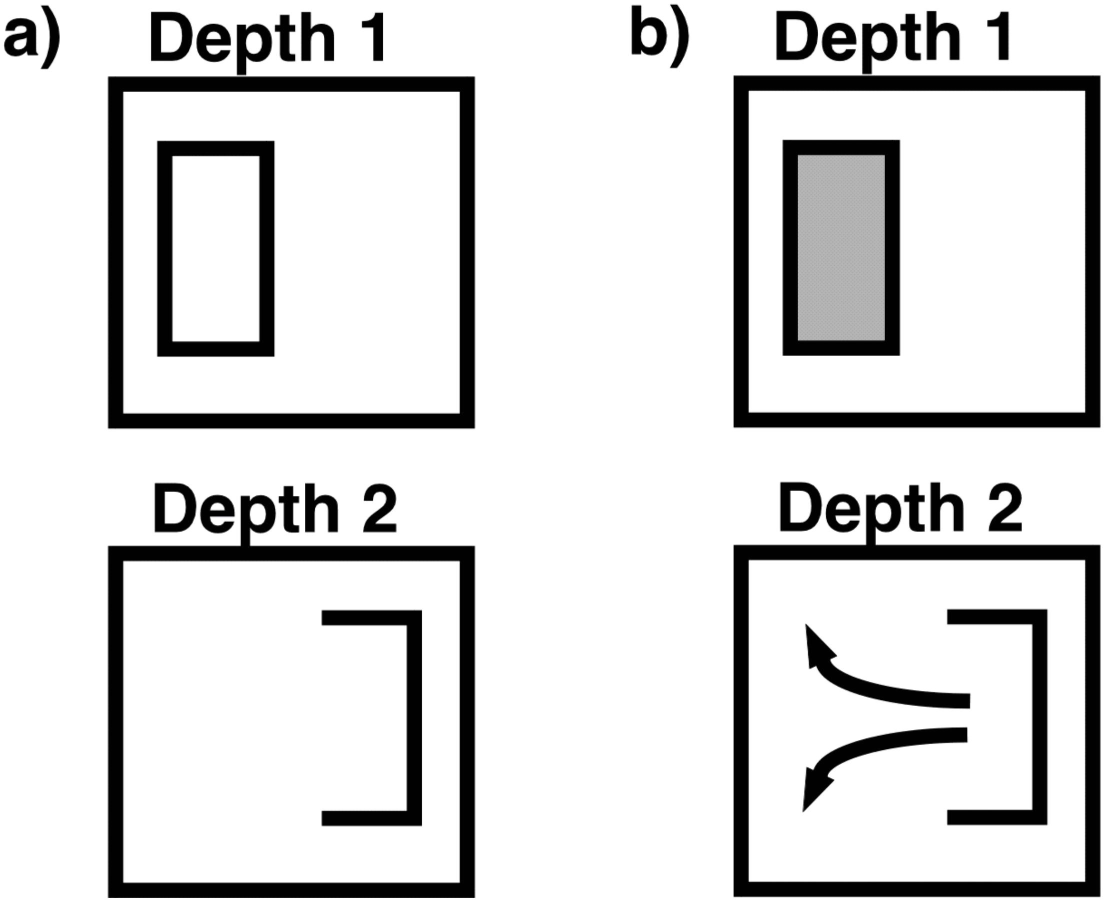


Figure 4. (a) Open and connected boundaries; (b) Filling-in of surface lightness is contained or not depending on the connectedness of the boundary. Note that the monocular boundaries (i.e., two horizontal boundaries and the right vertical boundary) have been added to all depth planes whereas the binocular 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 Cao and Grossberg (2005).]

(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 build boundaries of objects in front of textured backgrounds whose relative contrasts, where they contact the object, may reverse along the object's perimeter. Surface representations 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 via a process called 3D *surface capture*. The present model considers only the filling-in of achromatic lightness. See Grossberg and Hong (2006) for simulations of the filling-in of surface color.

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. The 3D sLAMINART model extends FACADE theory to spiking neurons interacting in laminar cortical circuitry, with a corresponding expansion in explanatory and predictive range. One of these boundary-surface interaction properties is that visible surfaces arise in cortical area V4 only if they are enclosed by *connected* boundaries. In particular, suppose, 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 contain the filling-in process, and thereby support a visible surface percept at the depth corresponding to the binocular boundary if all other constraints are satisfied. However, if the vertical binocular boundary is missing at a different depth plane, then the total boundary is not connected. A visible percept will thus not arise at that depth because lightness and color signals can flow out of the boundary gap and around the boundaries, thereby eliminating any surface contrast across space. This example illustrates how the Monocular-Binocular Interface Problem (constraint (4) above), and the Correspondence Problem (constraint (3) above), together influence properties of visible percepts of 3D surfaces.

(7) Surface-to-boundary feedback ensures perceptual consistency and initiates figure-ground separation. Boundaries and surfaces form according to computationally *complementary* rules. How do they nonetheless give rise to a percept whose perceived boundaries and surfaces are perceptually *consistent*? As proposed in Grossberg (1994), successfully filled-in regions within the surface representations send feedback to the boundary system (Figure 1). 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 a successfully 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, called *surface contours*, are realized by on-center off-surround networks that operate within disparity and across position within the surface system. The surface contour 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 farther 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. This boundary-enhancing property interacts with a developmental bias that favors the fixation plane to simulate stereopsis data that cannot otherwise be explained.

Surface contour signals also help to control where the eyes look to inspect and learn several different view categories whereby to recognize a given object's surface, and bind them into a learned view-invariant object category, before the eyes move to inspect and learn about different objects. The ARTSCAN model (Cao, Grossberg, and Markowitz, 2011; Fazl, Grossberg, and Mingolla, 2009; Grossberg, 2009) predicts how surface contour signals may generate predictive eye movement commands, via cortical area V3A, that determine where the eyes will look next, while helping to maintain spatial attention on an object that is being learned and recognized. Thus the 3D sLAMINART model is part of a larger emerging architecture for active vision wherein 3D boundary and surface representations help to control eye movements for searching, learning, and recognizing invariant object categories.

2. Model Description

As in the rate-based 3D LAMINART model (Cao and Grossberg, 2005), the spiking 3D sLAMINART model consists of two processing streams: a boundary stream and a surface stream (see Figure 1). The boundary stream runs from retina/LGN to V2 binocular boundaries via V1 monocular and binocular boundaries, while the surface stream runs from retina/LGN to V4 surfaces via V1 and V2 monocular surfaces. Figure 1a shows a laminar cortical circuit diagram of the 3D sLAMINART model, and Figure 1b shows a block diagram of the model. A mathematical description of the spiking model is provided in Section 4. As in the rate-based model, in order to reduce the computational load, the model currently simulates only horizontal and vertical contours and five depth planes. Five depth planes were chosen because they are enough to illustrate non-trivial depth separation. The model is extensible to any finite number of depth planes, where finer depth discrimination may be needed.

2.1. Integrate and fire neurons. Each neuron is modeled as an integrate-and-fire neuron. The neuron fires when its membrane potential reaches a pre-defined firing threshold, and then its membrane potential is reset to the rest state immediately after firing. See equations (1-6) in Section 4 for more details.

2.2. LGN cells. 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. See equations (7-9) for more details.

2.3. V1 monocular simple cells. The V1 monocular boundaries are computed from V1 monocular simple cells. 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. Off-surround interactions in layer 4 suppress noise and sharpen the tuning curves of the final simple cell responses. See equations (10-16) for how these spiking integrate-and-fire equations are defined.

2.4. V1 binocular simple cells. Layer 3B of V1 begins the process of stereoscopic fusion at the binocular simple cells in layer 3B of the interblobs. As described in Section 2.3, inputs from the LGN activate monocular simple cells in layer 4 of the V1 interblob regions. Left and right eye monocular simple cells then activate binocular simple cells in layer 3B. The depth sensitivity of binocular simple cells is determined by the 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 binocular simple cell in layer 3B (Figure 2a). This enables the *same-sign hypothesis* to be implemented, since layer 3B simple cells are selective for both binocular disparity and contrast polarity.

The binocular simple cells in layer 3B achieve stereoscopic fusion with the help of 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.

How do these inhibitory interneurons ensure that binocular simple cells respond only to a narrow range of disparities? As in the definition of the bipole grouping property (see Sections 1.3(5) and 2.7), the inhibitory interneurons inhibit each other through recurrent inhibitory interneurons, as well as their target binocular simple cell. The recurrent inhibitory interactions normalize the total activity of the interneuronal population, thereby ensuring a "two-against-one" computation that ensures disparity selectivity. The circuits for generating V1 binocular simple cells and V2 binocular bipole cells thus seem to be computationally homologous. It will be interesting to discover in future studies how this homology may arise during brain evolution and development.

It is challenging to implement the obligate property using spiking integrate-and-fire neurons. In particular, spikes from the left and right eye inputs do not always arrive at the same time. See equations (17-19) for how these spiking neuron equations are defined. Cao and Grossberg (2011) provide a detailed analysis of how the stereoscopic fusion property and other desired properties can be achieved using these spiking neuron equations.

2.5. V1 layer 2/3 complex cells. 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 can thus respond all along an object's boundaries even if its contrast polarity, with respect to the background, reverses as its boundary is transversed (Figure 2b). Layer 2/3 implements contrast-invariant boundary detection. See equations (20-21) for more details.

2.6. V2 layer 4 binocular cells. Monocular and binocular inputs are combined in layer 4 of the V2 pale stripes. In particular, V2 layer 4 cells receive inputs from V1 layer 2/3 left and right monocular complex cells and binocular complex cells. 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 (Figure 4; see Yazdanbakhsh and Watanabe, 2004, for consistent data.). The layer 4 cells also receive surface contour feedback signals from left and right V2 monocular surfaces that are formed in the V2 thin stripe region; see Section 1.3(7). These surface-to-boundary feedback inputs modulate the corresponding V2 layer 4 cells: 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. As shown in our simulations, these surface-to-boundary feedback signals play an indispensable role in explaining percepts of some stereo displays, such as da Vinci stereopsis. See equation (22) for more details.

2.7. V2 layer 2/3 complex cells. The binocular cells in V2 layer 2/3 receive inputs from layer 4 cells. Layer 2/3 implements 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 that inhibit themselves as well as their target complex cells. This balance of excitation and self-normalizing inhibition achieves the "two-against-one" bipole grouping property (Grossberg, 1999a, 1999b; Grossberg, Mingolla 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 (Figure 2b).

Achieving perceptual grouping using spiking integrate-and-fire neurons is a major accomplishment of the brain. In rate-based models of grouping, synchronous arrival of inputs from both sides of the neuron's receptive field is easily achieved. In a network of spiking neurons, synchrony of individual spikes is not guaranteed. In addition, the collinear excitation in the grouping process can drive layer 2/3 complex cell firing excessively unless a spike refractory period limits the maximum spiking frequency. See equations (23-29) for details.

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 if it has the same contrast polarity and approximately the same magnitude of contrast. As a result, there is a Correspondence Problem because false matches may occur in V1. Such false matches are suppressed in V2 via a disparity filter, which is realized as part of the inhibitory interactions that control perceptual grouping by long-range horizontal connections in V2 layer 2/3. The bipole grouping process hereby combines suppression of false matches with long-range Gestalt grouping processes. See equation (30) for details.

2.8. V2 monocular surfaces. The network that processes the V2 monocular surfaces is located in the model V2 thin stripes. 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 1a). It has elsewhere 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 and transparency (Grossberg, 1994; Grossberg and Mingolla, 1985a; Grossberg and Swaminathan, 2004; Grossberg and Yazdanbakhsh, 2005), and lightness perception (Grossberg and Hong, 2006; Grossberg and Kelly, 1999; Grossberg and Todorovic, 1988). As illustrated by Figure 4, a perceived surface representation can rise from filling-in only if it is enclosed by a connected boundary.

The LGN cells, which obey integrate-and-fire membrane equations and interact through on-center, off-surround circularly symmetric receptive fields, first discount the effects of spatially non-uniform illumination; i.e., "discount the illuminant". The excitatory and inhibitory components of these receptive fields are balanced so that cell responses are attenuated in response to spatially uniform or slowly varying stimulation. Model LGN cells hereby respond preferentially to luminance borders. At the later filling-in stage, these illuminant-discounted surface-border signals propagate throughout surface regions that are completely enclosed by connected boundaries to complete the lightness representation. The filling-in process has often been modeled by a boundary-gated diffusion equation (Grossberg and Todorovic, 1988). 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 from V2 layer 2/3 that are received at the monocular surfaces create such barriers and, together with the lightness information received from LGN, can fill-in monocular surface representations in V2 thin stripes.

To model the surface filling-in process using spiking neurons, one challenge is how to implement the boundary-gated filling-in coefficients. In the rate-based model, these coefficients are defined by the weighted inverses of boundary strengths, which are the equilibrium solutions of a shunting competitive interaction (Grossberg and Todorovic, 1988). This is impracticable for spiking neuron models in which boundary signals are discrete spikes. A mechanism is needed to generate boundary gating signals that persist through time. A differential equation, rather than an equilibrium solution, now describes the continuous inactivation, or habituation, of the boundary gate through time. Equations (31-38) describe the spiking dynamics of boundary gating and monocular surface filling-in.

2.9. Surface-to-boundary feedback signals. Successfully filled-in monocular surfaces send contour-sensitive surface-to-boundary feedback signals into V2 layer 4 (Figure 1b); see Section 1.3(7). These surface-to-boundary surface contour signals modulate the activities of V2 boundary cells so that the boundaries that surround the successfully filled-in surfaces are enhanced and redundant boundaries are suppressed. See equations (39-42) for details.

2.10. V4 surfaces. Although monocular surfaces are formed in V2 thin stripes, they are predicted to be invisible, or amodal, and thus do not subserve visible 3D surface percepts. The visible 3D surface percepts are proposed to be generated in cortical area V4, where 3D figure-ground separation of object surfaces is also predicted to be completed. 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 2.8, except V4 combines monocular lightness signals from both eyes. See equations (43-45) for details.

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 in the rate-based model (Cao and Grossberg, 2005), 18 challenging psychophysical experiments are simulated. Like the model diagram shown in Figure 1a, each simulation figure should be read from the bottom up, with the bottom three rows representing the input, LGN and the V1 boundary representations, the next two rows representing the V2 boundary representations, and the top row representing the V4 surface representations. Furthermore, 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. Each eye's stimulus was presented on a grid 30 units high and 60 units wide except Figure 15, which was presented on a 30 by 70 grid, and Figure 19, 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 5-22, the light gray bars (if any) had a luminance of 1 and the dark gray bars 0.1. In Figures 23-25, white bars were represented by a luminance of 4, and black by 0.1. Simulations were performed using the *Matlab* software package. A total of 2000 iterations are run for each experiment with time step 0.001. In the simulation figures, each plot shows the firing rates of cells. The firing rates are computed from counting all spikes from iteration 1 to 2000 for the first four rows that represent LGN to V2 layer 4, and counting spikes from iteration 500 to 2000 for the top two rows (V2 layer 2/3 and V4) when the network approaches a stable state.

3.1. Variations of da Vinci Stereopsis. In the 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 first row of Figure 5. 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. The 3D sLAMINART model responds to this input as follows: The spiking rate of LGN cells (computed by counting spikes generated from equations 7-8) and V1 monocular simple cells (computed by counting all spikes generated from equations 14-15) are shown in the second row. Through stereoscopic fusion (equations 17-19) the boundaries of the thick bar are then registered binocularly in the near disparity plane in V1, as shown by the second plot of the third 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. 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 fourth row of this figure (equation 22). The spiking binocular boundary signals in the third row are also added to V2 layer 4, overlapping with the boundary signals of the thick bar in the second plot and with the rightmost boundary in the fourth plot.

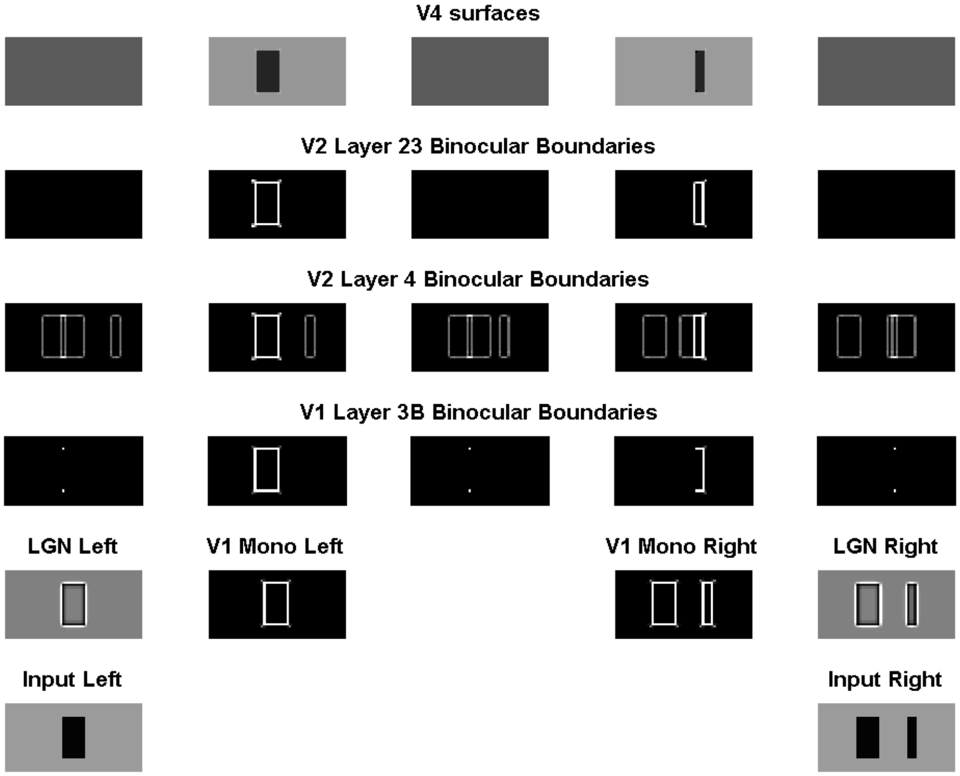


Figure 5. Simulation of the da Vinci stereopsis display reported experimentally by Nakayama and Shimojo (1990). The first row represents the inputs, the second row LGN (the outer two plots) and the V1 monocular boundaries (the middle two plots), the third row the V1 binocular boundaries, the fourth row the V2 layers 4 boundaries, the fifth row the V2 layer 2/3 final boundaries, 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. In other words, the plots from left to right, in turn, denote very near, near, fixation, far and very far depth planes. 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 gray background. All simulation plots use a similar format.

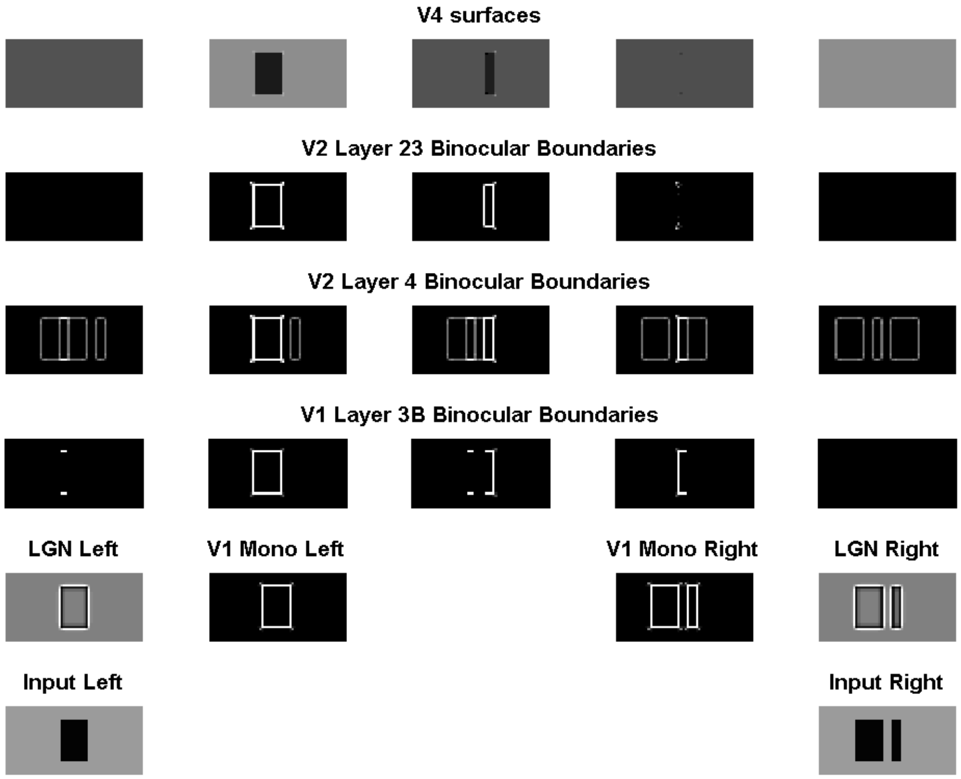


Figure 6. Simulation of a variation of the da Vinci stereopsis display of Figure 5. See text for details.

Spiking boundary signals in V2 layer 4 input into V2 layer 2/3. V2 layer 2/3 acts as the disparity filter to solve the Correspondence Problem while also controlling perceptual grouping by horizontal connections (equation 23). The spiking binocular boundary inputs in the second plot are much stronger, and fire target V2 layer 2/3 cells first. Once these cells fire, they use the disparity filter line-of-sight inhibition to prevent all other cells from firing that receive only monocular inputs and share their lines-of-sight. This property is slightly different from the rate-based model, in which cells with weaker inputs do get active initially, but are suppressed later by the disparity filter.

The rightmost boundary cells in the fourth plot receive stronger binocularly fused inputs and can thus fire. The left edge of the thin bar does not binocularly fuse with any other edges and hence is not in the lines-of-sight of any binocularly fused edges. As a result, V2 layer 2/3 cells in all depth planes corresponding to the left edge of the thin bar can fire initially, though at a later time than the cells receiving binocularly fused inputs. Once this has happened, then only the far depth plane has connected boundaries of the thin bar and thus can form a monocular surface via filling-in (equations 31-38). The filled-in surface then sends surface-to-boundary feedback signals (equation 39-42) to corresponding V2 layer cells and enhances their firing rates (equation 22). 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. 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. This happens so fast that the left boundary of the thin bar in the far depth cannot be depressed by the corresponding left boundary in the fixation plane via the fixation plane advantage. 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 5. 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).

Now let us 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 the far depth plane (see Figure 6). This case will happen by moving the thin bar of the right input closer to the thick bar than in the original case shown in Figure 5.

Similar to Figure 5, LGN cells and V1 monocular simple cells fire first, as shown in the second row. The boundaries of the thick bar are then 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 third row of Figure 6, respectively. 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 fourth row of Figure 6. The binocular boundaries in V1 (displayed in the third 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 spiking binocular boundary inputs in the second plot are much stronger, and fire target V2 layer 2/3 cells first.

Once these cells fire, the disparity filter line-of-sight inhibition prevents all other cells from firing that receive only monocular inputs and share their lines-of-sight. This includes the left edge of the thin bar in the fixation plane and its right edge in the far depth. Both the right edge of the thin bar in the fixation plane and its left edge in the far depth receive strong binocularly-fused inputs from layer 4, and can thus fire initially. But only the thick bar in the near depth can form a closed boundary contour so far and hence form a filled-in V2 monocular surface that can send surface-to-boundary feedback signals to V2 layer 4. The feedback loop makes the thick bar boundaries in the near depth become much stronger, and can hence depress the binocularly-fused boundary in the far depth. Once this happens, the left boundary of the thin bar in the fixation plane can fire because it is not depressed by the binocularly-fused boundary in the far depth any more. See Figure 7 for the spiking time course of this competition.

As a result, a closed boundary contour of the thin bar is formed in the fixation plane. Note that, although the right boundaries of the thick bar in the near depth and of the thin bar in the fixation plane are on the same line-of-sight, they cannot then depress each other since they are both firing strongly with surface-to-boundary feedback and the self-scaling process prevents either of them from firing excessively. The final V2 boundaries are shown in the fifth row.

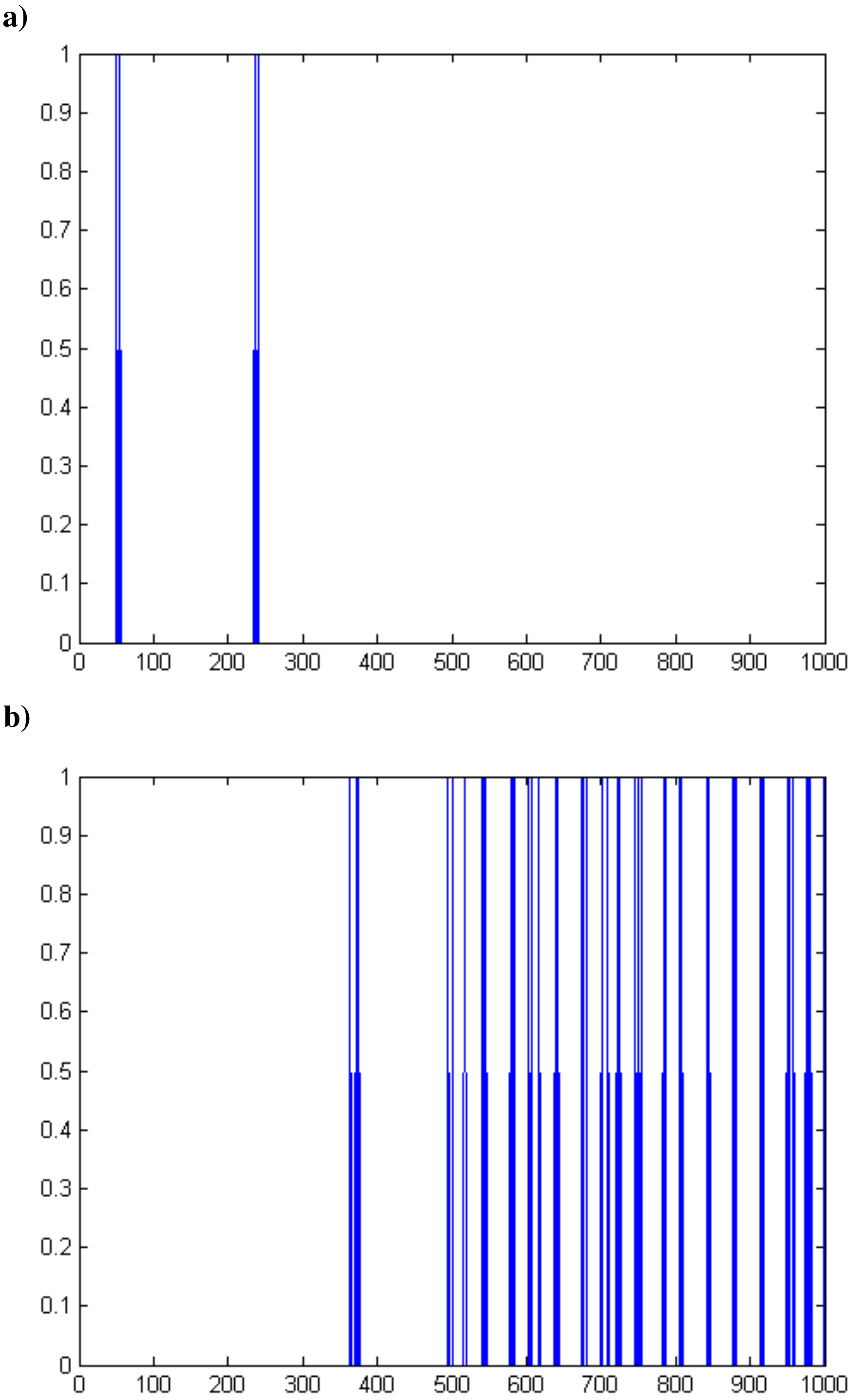


Figure 7. (a). Spiking dynamics of the V2 layer 2/3 cell in the far depth for the mid point in the left edge of the thin bar in Figure 6. The cell fires initially but is depressed after less than 300 iterations. Other cells in the far depth for the left edge have similar spiking dynamics. (b). Spiking dynamics of the V2 layer 2/3 cell in the fixation plane for the mid point in the left edge of the thin bar. The cell is depressed initially but after the cell in (a) is depressed it can then fire later. Other cells in the fixation plane for the left edge have similar spiking dynamics.

V4 then 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 by Nakayama and Shimojo (1990).

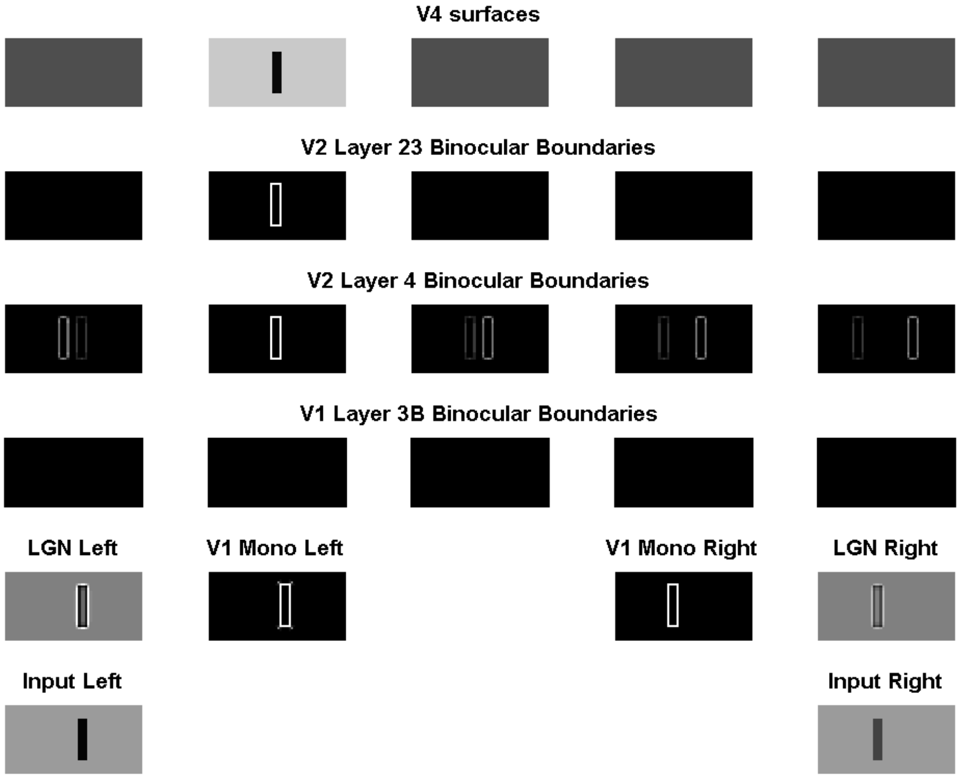


Figure 8. Simulation of the basic dichoptic masking studied by McKee et al. (1994). See text for details.

3.2. Dichoptic masking. In the experiments of 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 bottom row of Figure 8. 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 fourth row of Figure 8. Note that the left and right monocular boundaries coincide in the near depth plane represented by the second plot of this row, and 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. Since the boundary inputs in the near depth are much stronger than the inputs in other depths, the V2 layer 2/3 boundary cells in the near depth fire first and then depress all boundaries in other depths. The final V2 boundary representations are shown in the fifth row. 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).

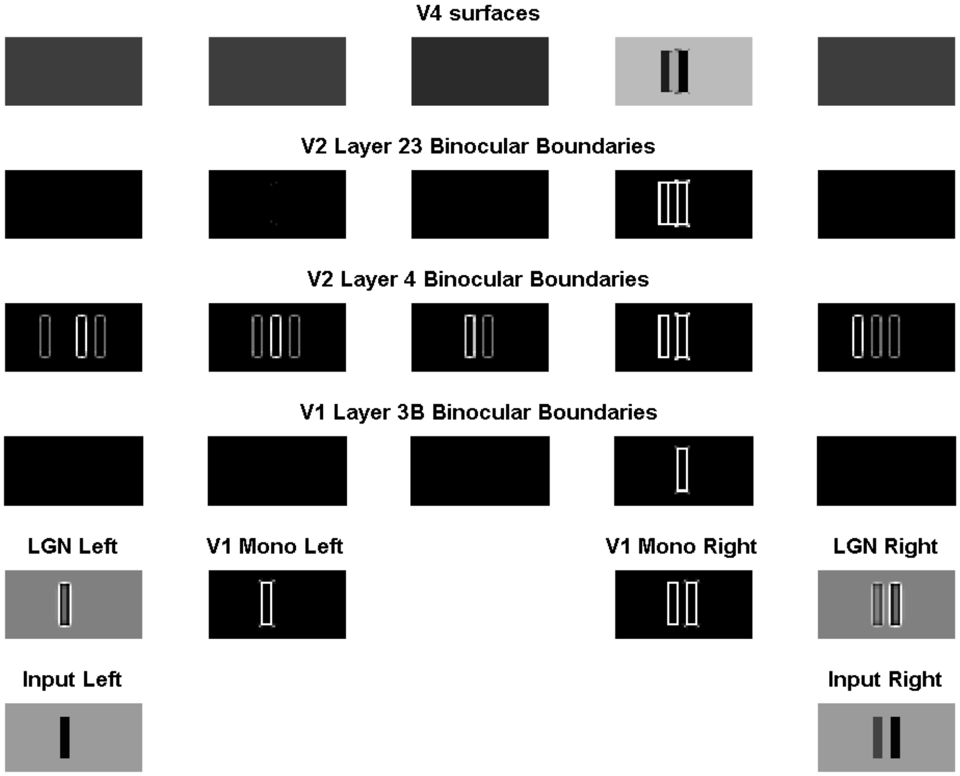


Figure 9. Simulation of the release from dichoptic masking studied by McKee et al. (1995). See text for details.

Figure 9 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 explanation of the ensuring percept is as follows: The boundaries of the high contrast bar are registered binocularly in the far depth plane in V1, as shown in the fourth plot of the third row. The low contrast bar in the right input does not binocularly fuse with the high contrast bar in the left input since their contrasts differ too much. Monocular and binocular boundaries are added together in V2 layer 4, as shown in the fourth row. Cells in V2 layer 4 then input into V2 layer 2/3. Layer 2/3 boundary cells of right bar in the far depth (the fourth plot) fire first since they receive much stronger binocularly-fused inputs. These cells then depress all other cells which share their lines-of-sight, including cells corresponding to the both bars in the fixation plane.

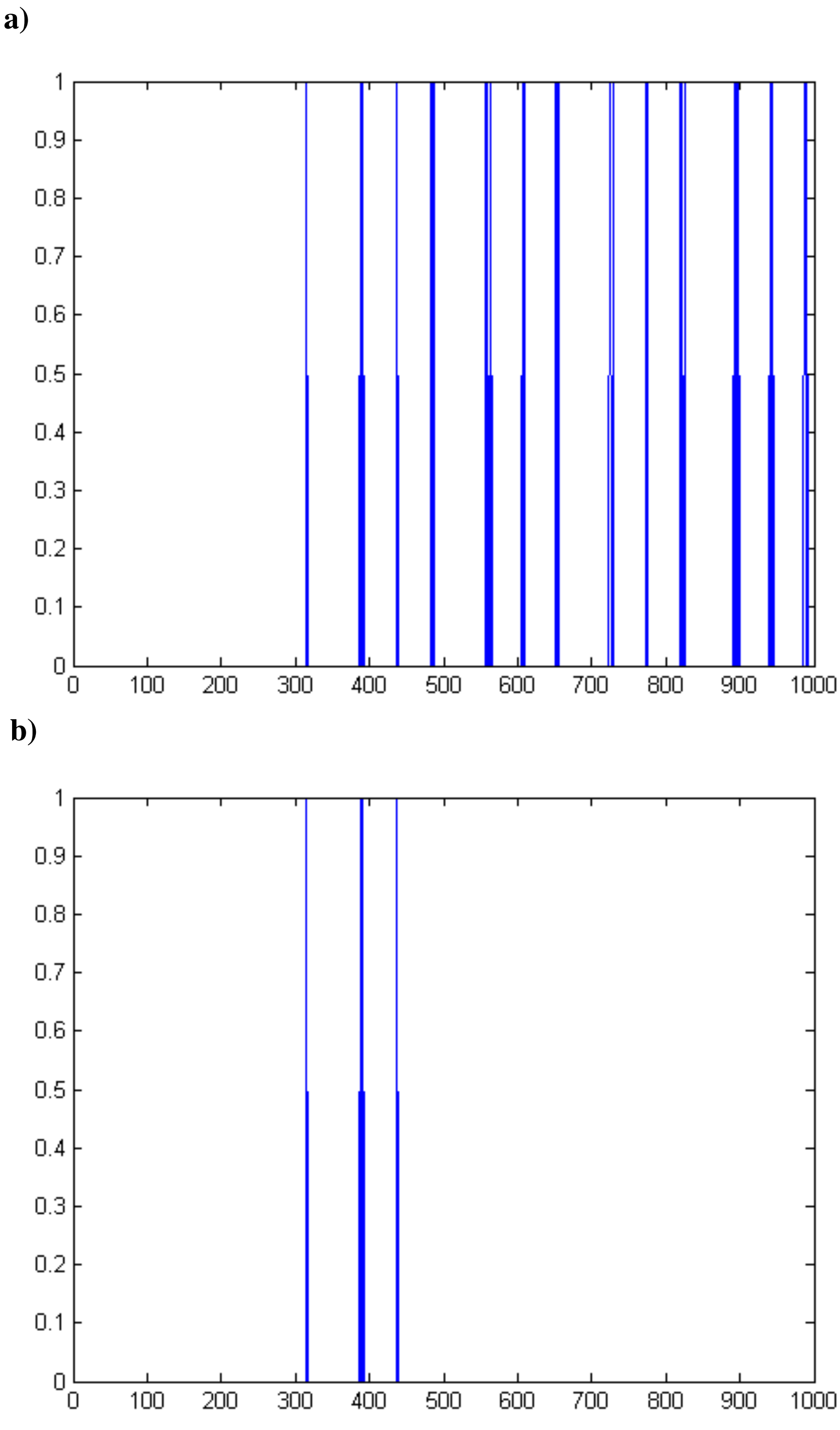


Figure 10. (a). Spiking dynamics of the V2 layer 2/3 cell in the far depth for the mid point in the left edge of the low contrast bar in Figure 9. The cell fires after about 300 iterations. Other cells in the far depth for the left edge have similar spiking dynamics. (b). Spiking dynamics of the V2 layer 2/3 cell in the near depth for the mid point in the left edge of the low contrast bar. The cell fires at the same time as the cell in (a) but is depressed later. Other cells in the fixation plane for the left edge have similar spiking dynamics.

The cells corresponding to the low contrast bar in depth planes other than the fixation plane are then free to fire. Since the low contrast bar boundary cells in the near and far depth planes receive the same V1 monocular inputs, they cannot depress each other by themselves. But the long-range perceptual grouping (equation 23), together with surface-to-boundary feedback, makes the boundary cells of both bars in the far depth firing much stronger. As a result, the boundary cells in the near depth are depressed. See Figure 10 for a spiking time course of this competition process. The final V2 boundary representations are shown in the fifth row of Figure 9. V4 then fills in surfaces in those regions that are completely enclosed by a connected boundary. This produces a percept of both bars in the far depth, represented by the fourth plot of the top row. The model therefore correctly predicts that the low contrast bar is released from masking by the high contrast bar, as reported experimentally by McKee et al. (1995).

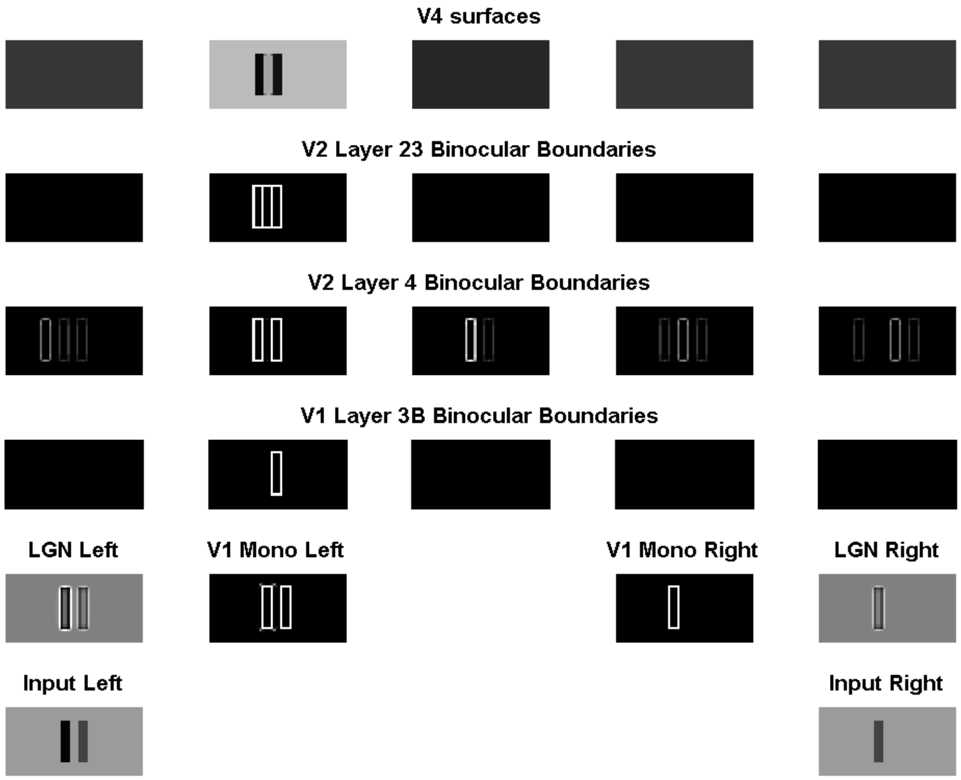


Figure 11. Simulation of a variation of the release from dichoptic masking studied by McKee et al. (1995). See text for details.

Based on the same reasoning, the spiking model can explain other dichoptic masking cases. The simulations that demonstrate this claim are shown in Figures 11-13. Figure 11 shows 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 correctly predicts that the low contrast bar is released from masking by the high contrast bar.

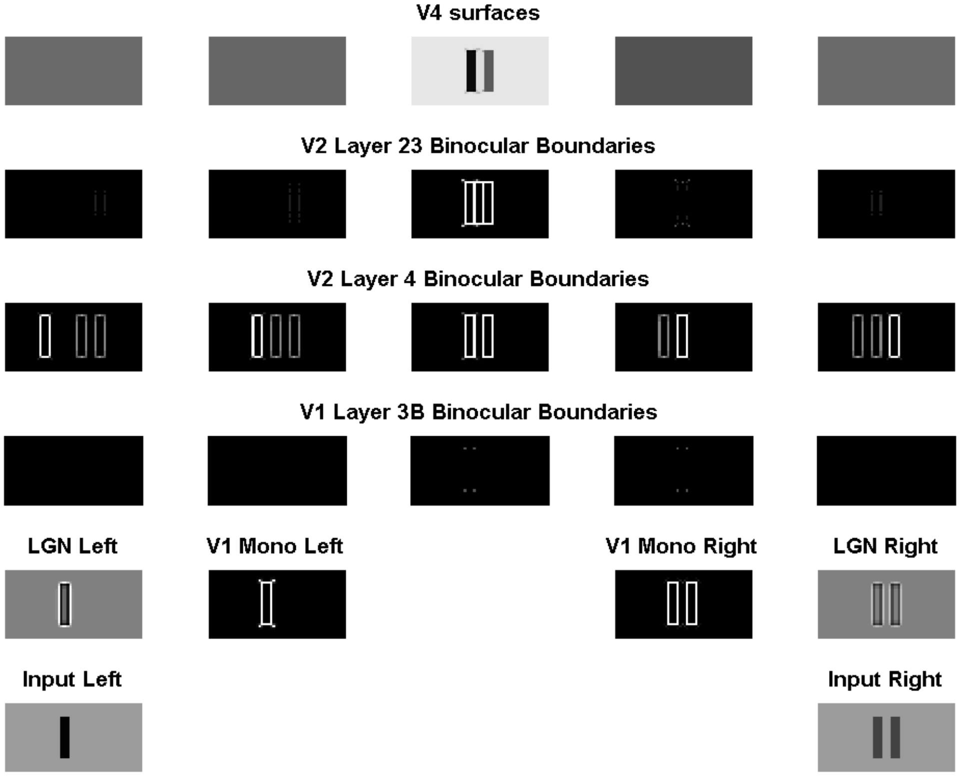


Figure 12. Simulation of the return to dichoptic masking studied by McKee et al. (1995). See text for details.

Figure 12 shows 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 13 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 simultaneously fused with the two right input bar and so masks them both equally, as reported experimentally by McKee et al. (1995).

3.3. 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 14. The model 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 third 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 fourth 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

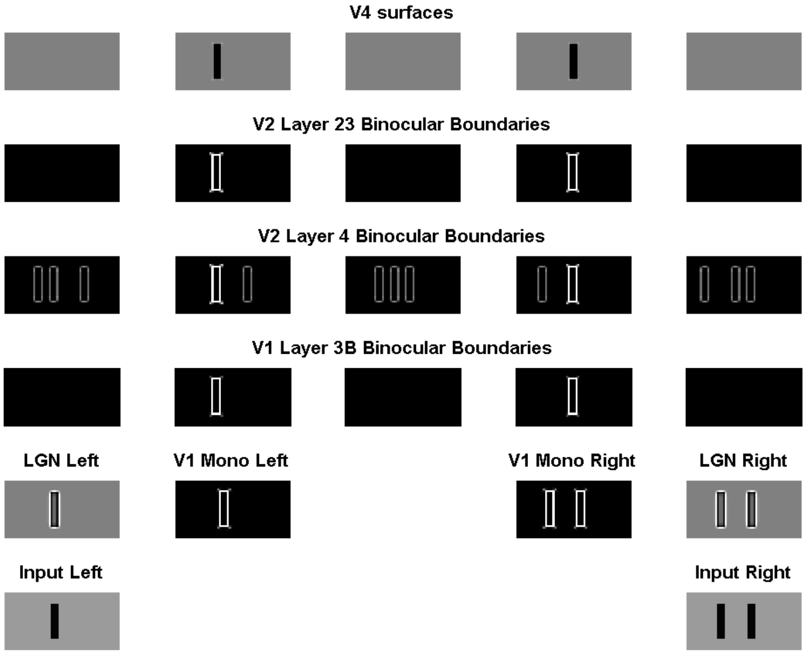


Figure 13. Simulation of dichoptic masking in Panum's limiting case studied by McKee et al. (1995). See text for details.

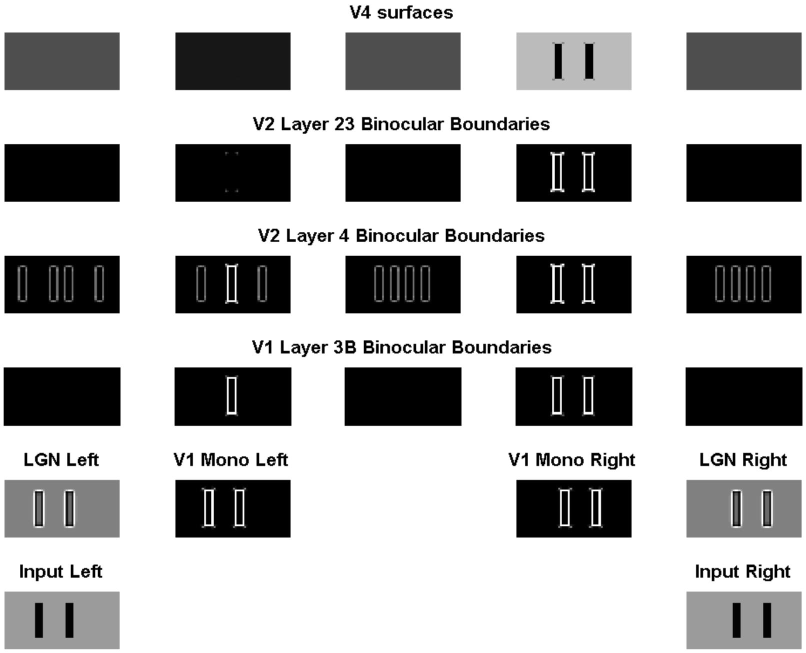


Figure 14. Simulation of the control experiment of the correspondence problem studied by Smallman and McKee (1995). See text for details.

the fourth plot. All boundaries in V2 layer 4 then input into V2 layer 2/3. The stronger boundary cells quickly fire and depress all other cells 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 14. The model correctly predicts that subjects see both bars in the far depth plane, as shown in the fifth row.

In a more complicated version, either eye was presented with three identical bars with the same high contrast. The simulation is shown in Figure 15. 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, as shown in the bottom row of Figure 16. 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 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 third row of Figure 16. 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 fourth 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 boundary cells which receive binocular input from V1, being stronger, fire first and depress all other boundary cells that share their lines-of-sight. 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 depress boundary cells of the lower contrast bar in all other depth planes, because these cells do not share any of their lines-of-sight. Since the fixation plane is favored, with additional support from perceptual grouping, the boundary cells of the lower contrast bar in the fixation plane can quickly depress its boundary cells in other depth planes via disparity filter line-of-sight inhibition. Figure 17 shows the spiking time course of this competition process. The final V2 boundary representations are shown in the plots of the fifth row of Figure 16. 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 18. The explanation is similar and therefore omitted.

3.4. 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 19, which shows the perceived Venetian blind percept in the top row of Figure 19. The spiking model explains this case along the lines proposed by Cao and Grossberg (2005), and thus the detailed explanation is omitted.

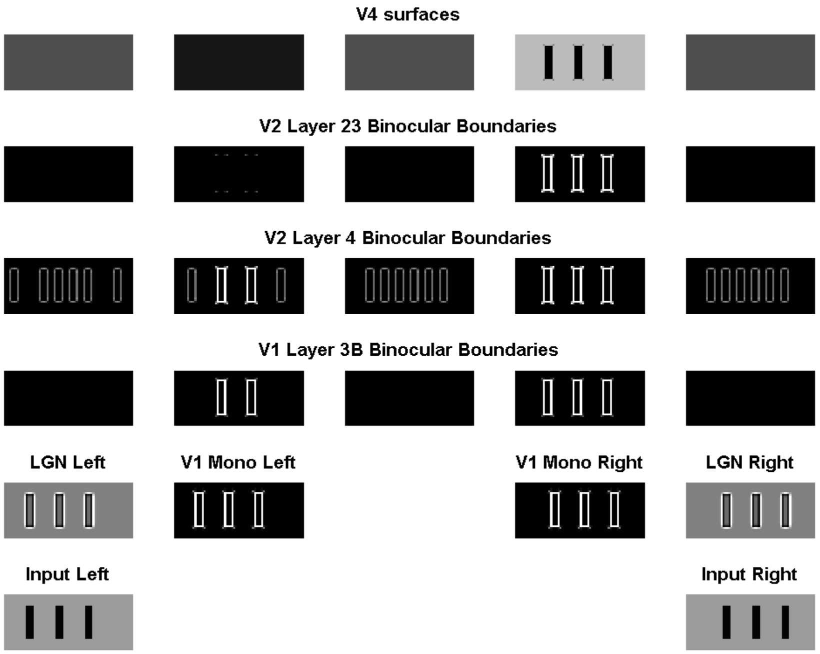


Figure 15. Simulation of a more complicated variation of Fig.14. See text for details.

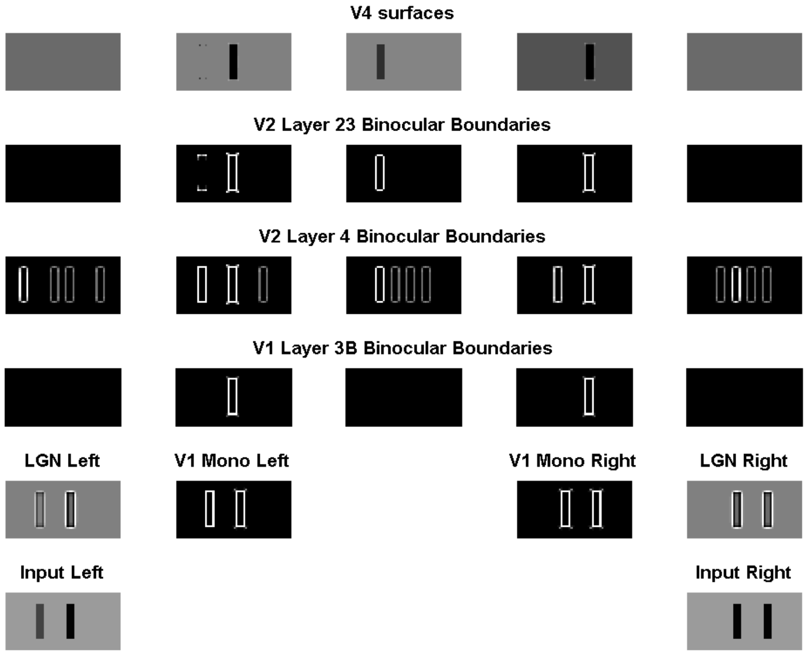


Figure 16. Simulation of a contrast variation of the correspondence problem reported by Smallman and McKee (1995). See text for details.

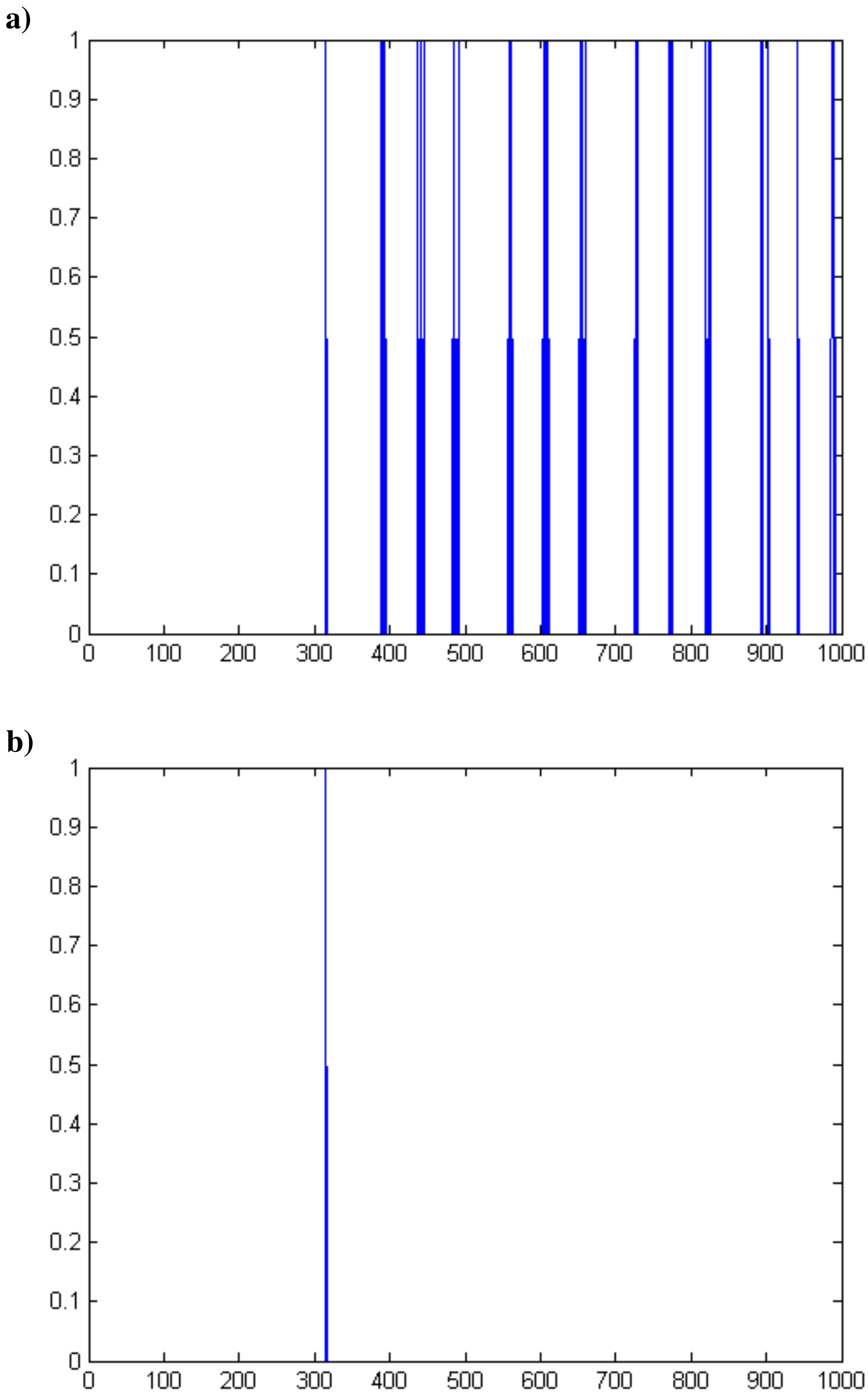


Figure 17. (a). Spiking dynamics of the V2 layer 2/3 cell in the fixation plane for the mid point in the right edge of the low contrast bar in Figure 16. The cell fires after about 300 iterations. Other cells in the fixation plane for the edge have similar spiking dynamics. (b). Spiking dynamics of the V2 layer 2/3 cell in the near depth for the mid point in the right edge of the low contrast bar. The cell fires at the same time as the cell in (a) but is then depressed. Other cells in the fixation plane for the edge have similar spiking dynamics.

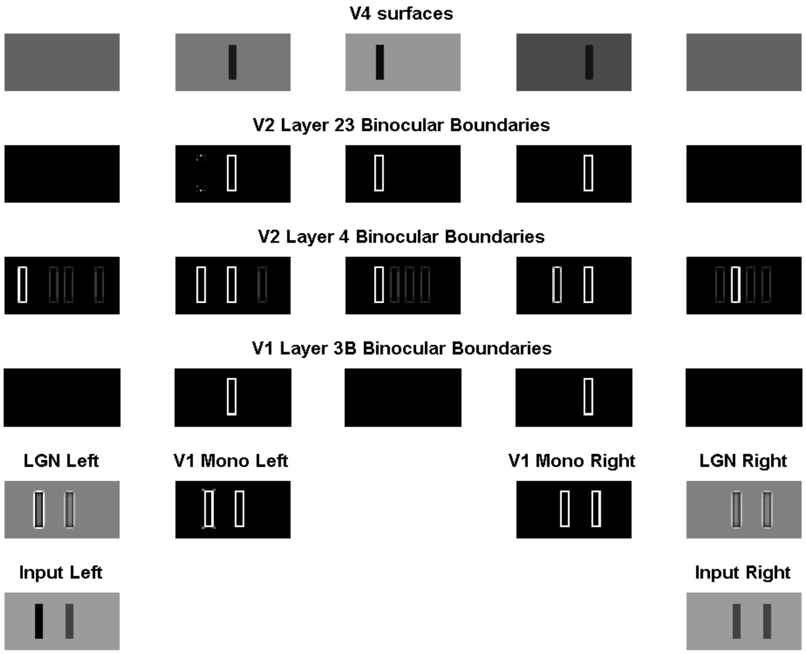


Figure 18. Simulation of another contrast variation of the correspondence problem reported by Smallman and McKee (1995). See text for details.

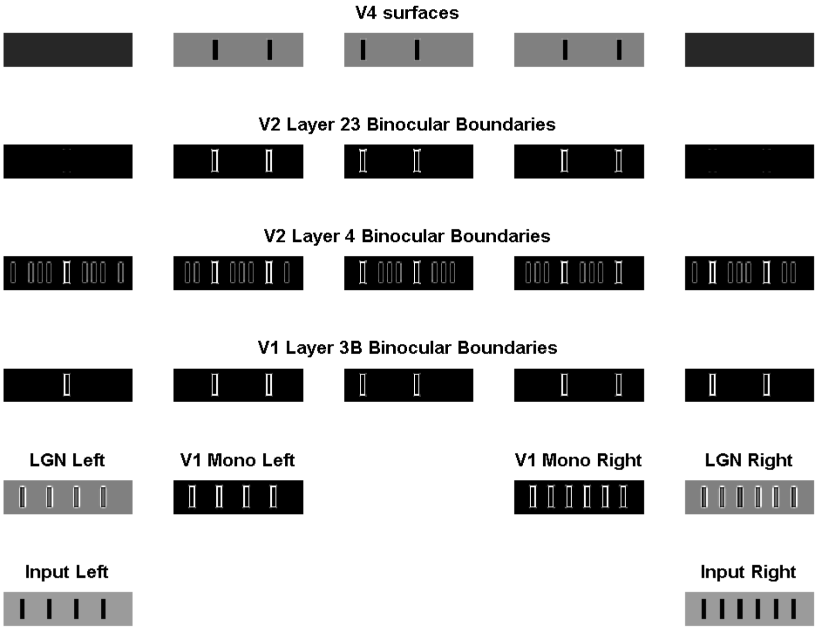


Figure 19. Simulation of the Venetian blind effect in Howard and Rogers (1995). See text for details.

3.5. da Vinci stereopsis experiments of Gillam, Blackbum, and Nakayama (1999). As discussed in Section 3.1, the spiking model correctly predicts variations of da Vinci stereopsis in Nakayama and Shimojo (1990). Now we consider the da Vinci stereopsis experiments in Gillam et al. (1999). In Figure 20, the right eye sees two thin bars and the left eye a single thick bar. The model correctly predicts that seeing two thin bars, the left in the near depth plane and other in the far depth plane, as reported experimentally by Gillam et al. (1999). Figure 21 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 again correctly predicts that three surfaces will be seen, each at a different depth, as reported experimentally by Gillam et al. (1999). The explanations for the two cases are very similar to those in Cao and Grossberg (2005), and thus are omitted.

3.6. Perceptual Closure. In the perceptual closure case, which was introduced by Cao and Grossberg (2005), 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 bottom row of Figure 22. 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. The spiking model produces the correct percept of a rectangular frame in a near disparity plane and a single bar in the fixation plane, as shown in the top row of Figure 22. The explanation is again similar to that given in Cao and Grossberg (2005) and thus omitted.

3.7. Stereopsis with opposite-contrast stimuli. A polarity-reversed stereogram studied by Grossberg and Howe (2003) is shown in the bottom row of Figure 23, 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 spiking model makes the correct prediction as shown in the top row of Figure 23.

A variation of the above polarity-reversed stereogram is shown in the bottom row of Figure 24, 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 (cf., Grossberg and Howe, 2003). The model again makes the correct prediction, as shown in Figure 24. The explanations to the two cases are again similar to those given by the rate-based model and 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 25, where the binocularly viewed thick bar and monocularly viewed thin bar had opposite luminance polarities. The model correctly predicts that perceiving a white thick bar in the near depth and a black thin bar in the far depth, as shown in the top row of Figure 25. The explanation is the same as the original da Vinci stereopsis case in Figure 5, 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.

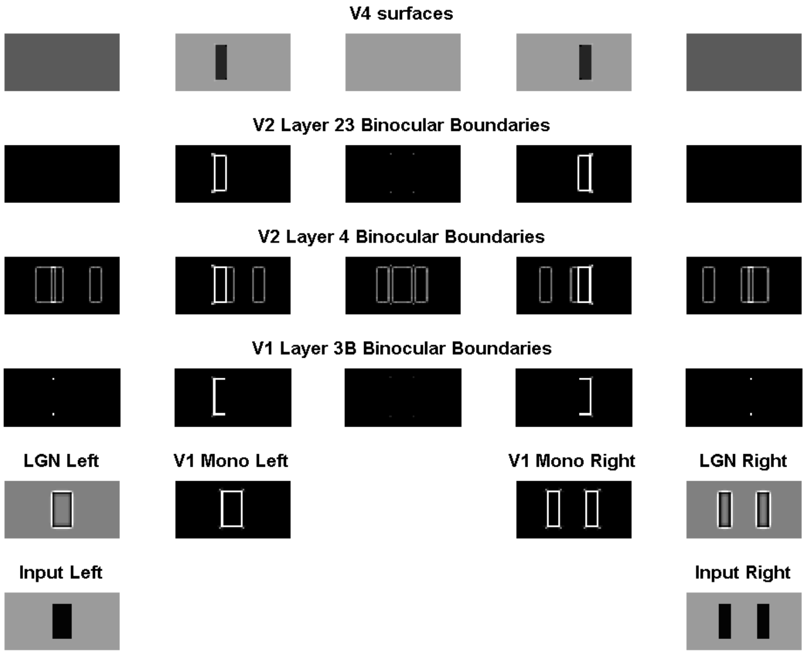


Figure 20. Simulation of the da Vinci stereopsis studied by Gillam et al. (1999). See text for details.

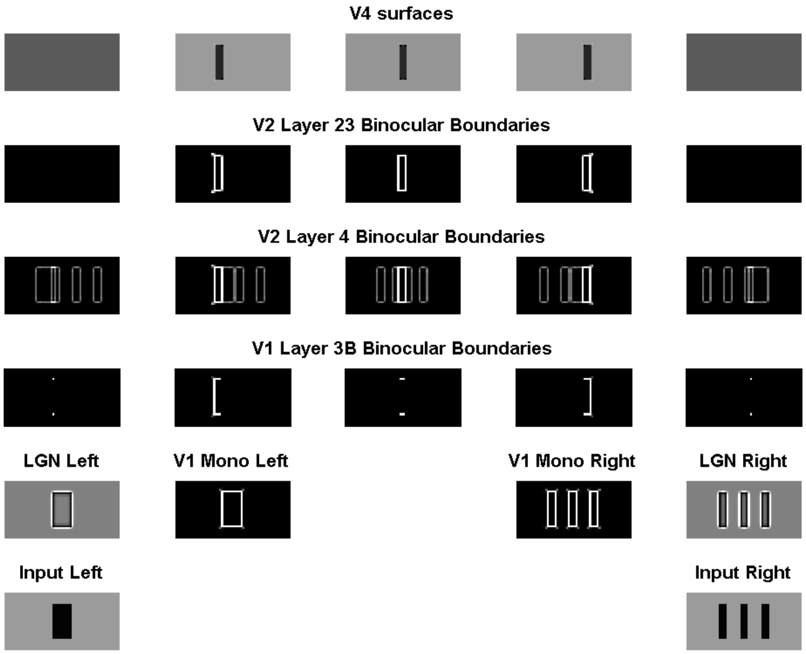


Figure 21. Simulation of a variation of the original da Vinci stereopsis of Figure 20. See text for details.

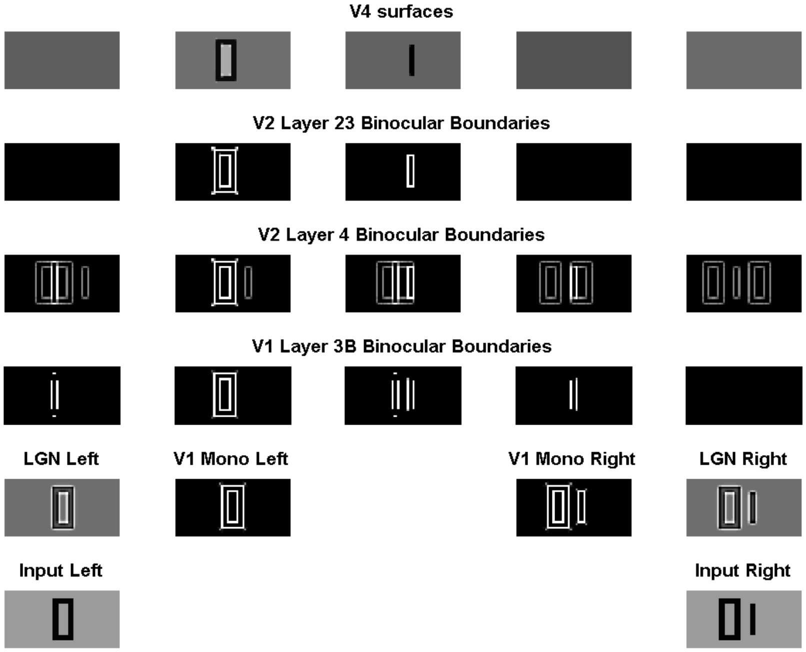


Figure 22. Simulation of perceptual closure case. See text for details.

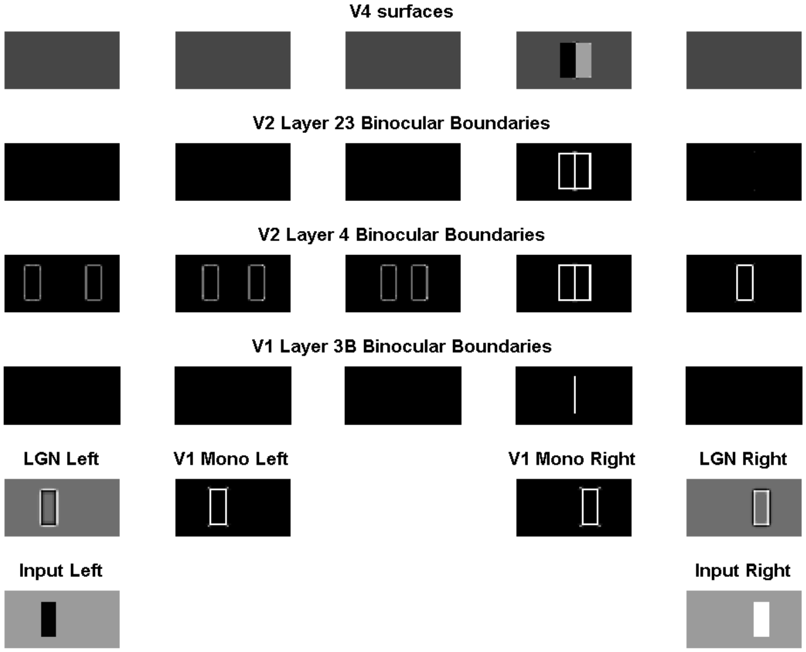


Figure 23. Simulation of stereopsis with a polarity-reversed stereogram. See text for details.

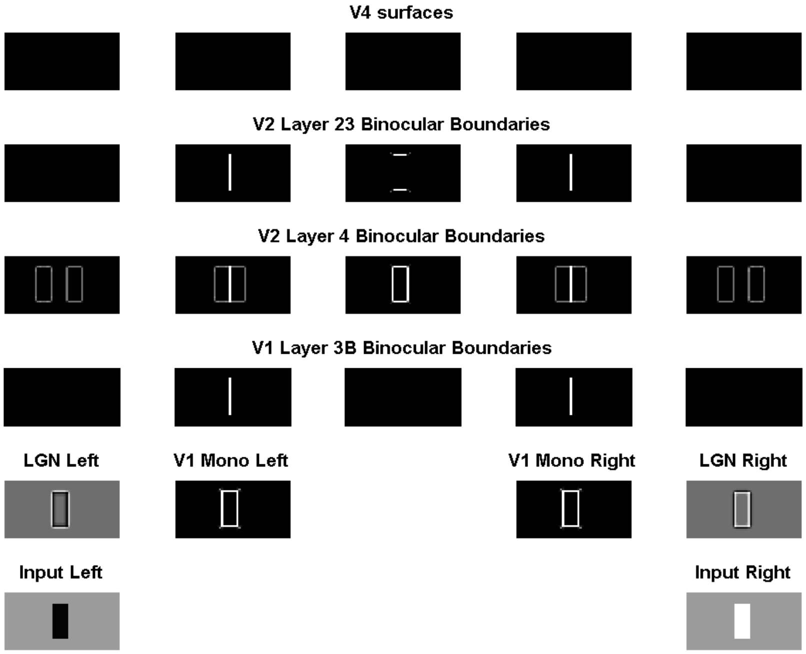


Figure 24. Simulation of a variation of Figure 23. See text for details.

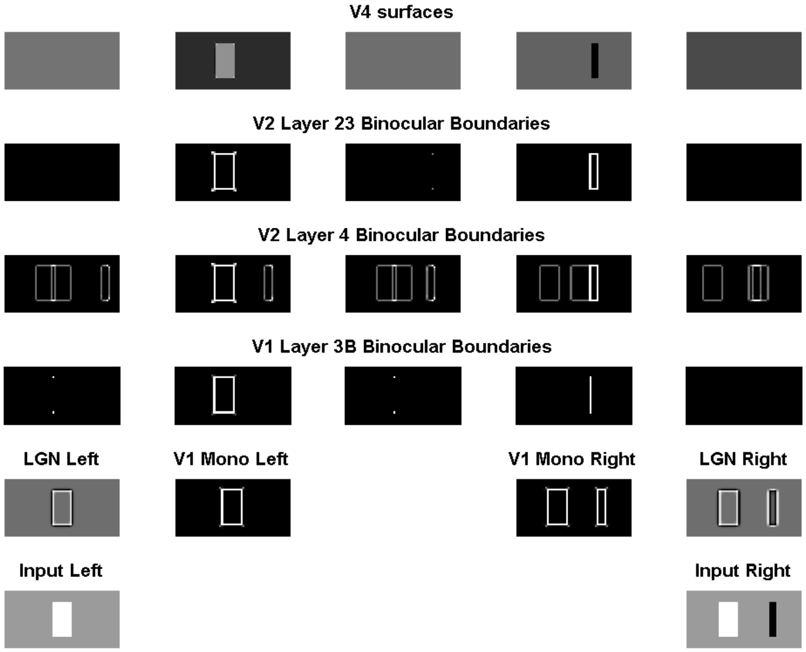


Figure 25. Simulation of the polarity-reversed da Vinci stereopsis studied by Nakayama and Simojo (1990). See text for details.

4. Model Equations

Integrate-and-fire neurons

Each neuron is typically modeled as a conductance-based integrate-and-fire neuron whose membrane potential, $v(t)$, is defined by the shunting equation:

$$\tau \frac{dv(t)}{dt} = (E_{leak} - v(t))g_{leak} + (E_{excit} - v(t))g_{excit}(t) + (E_{inhib} - v(t))g_{inhib}(t), \tag{1}$$

where τ is a capacitance constant, 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. The neuron fires and reset its membrane potential v to E_{leak} whenever v reaches up to a firing threshold v_{fire}. That is:

$$v = E_{leak}\text{, when } v \geq v_{fire}. \tag{2}$$

The above equations can be rewritten to make the reset state equal zero:

$$\begin{aligned} \tau \frac{dx}{dt} &= -Ax + g_{excit}(B - x) - g_{inhib}(C + x), \\ x &= 0, \quad when\ x \geq x_{fire}, \end{aligned} \tag{3}$$

where $x = v - E_{leak}$, $A = g_{leak}$, $B = E_{excit} - E_{leak}$, $C = E_{leak} - E_{inhib}$, and $x_{fire} = v_{fire} - E_{leak}$. Note that equation (3) can be further simplified by dividing both sides by B to make the maximum membrane potential equal to 1.

In some cases, the following simplified additive equation was also used:

$$\begin{aligned} \tau \frac{dx}{dt} &= -Ax + g_{excit} - g_{inhib}, \\ x &= 0, \quad when\ x \geq x_{fire}. \end{aligned} \tag{4}$$

A spike can be modeled as either binary or exponential decay type. A binary spike is a Dirac delta function $\delta(t)$. Its integral equals to 1 on any sufficiently small interval that contains the spike. It is zero when there is no spike. In the discrete-time simulation, a binary spike $\delta(t)$ is approximated as:

$$\delta(t) = \begin{cases} 1/\Delta t, & t_f \leq t < t_f + \Delta t, \\ 0, & otherwise, \end{cases} \tag{5}$$

where t_f is the firing time and Δt is the small time step. In the simulation, we have used time step $\Delta t = 0.001$. An exponential decay type spike $e(t)$ is defined as:

$$e(t) = \alpha e^{-\alpha(t - t_f)}, \tag{6}$$

where α is a decay rate and t_f is the last firing time. In the simulation, $\alpha = 500$ except otherwise noticed. In both cases, the integral (total energy) over a spike is 1. Figure 26 shows a spiking train and its exponential decay representation. In the following, an alphabetic letter (e.g., x) denotes the membrane potential of a neuron, and a hat of that letter (e.g., $\hat{x}$) denotes the spiking signal generated by that neuron. Whenever the membrane potential of a neuron reaches a given firing threshold, it generates a spike and is reset to zero. Except otherwise noticed, the spike is binary type due to its simplicity, and the firing threshold is set as 0.2.

a)

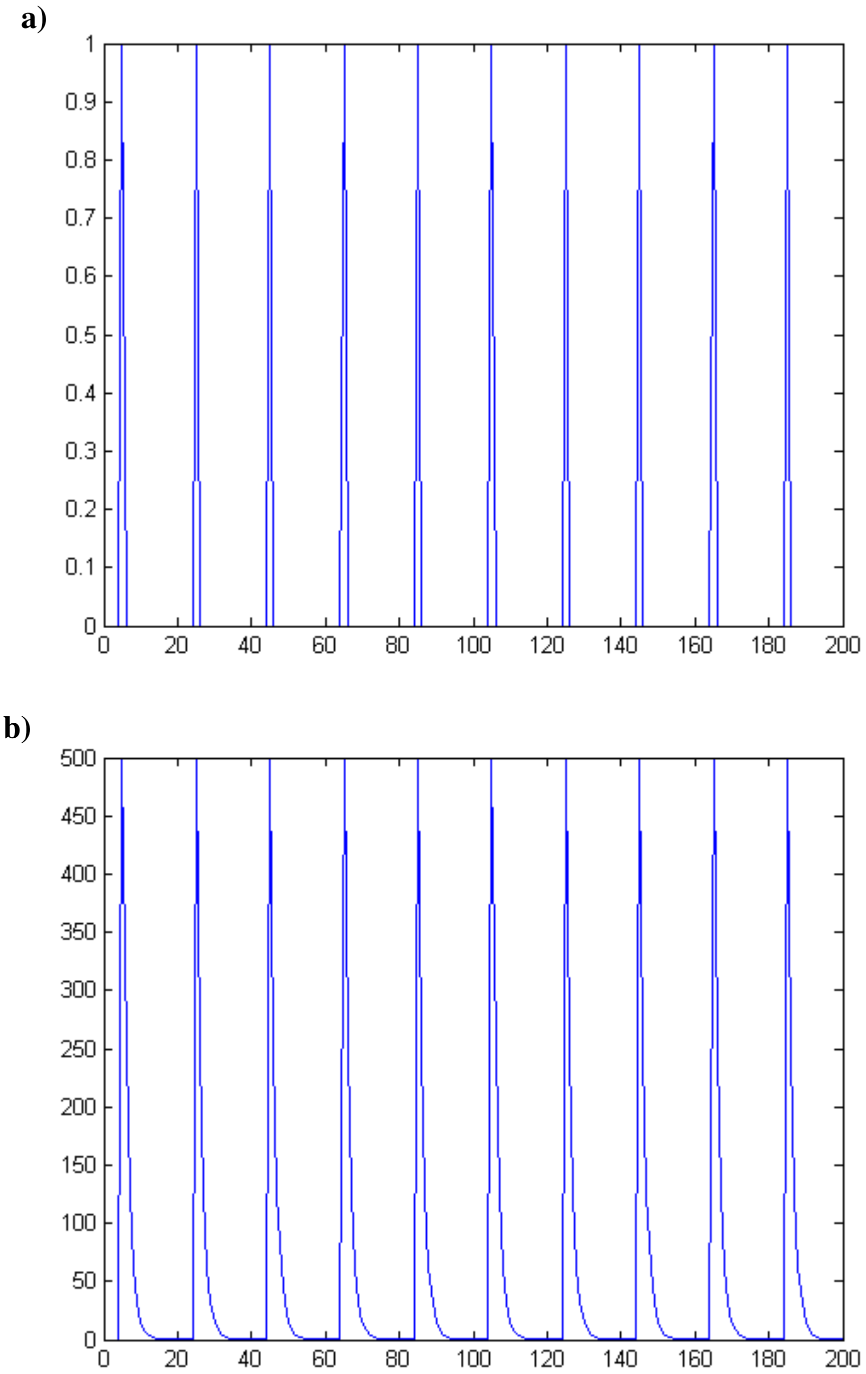


Figure 26. a). A spiking train. b). The exponential decay representation of a) with $\alpha = 500$ in equation (6).

LGN

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 LGN ON and OFF cell membrane potentials, x_{ij}, obey the following differential equation:

$$\tau \frac{dx_{ij}^{L/R,+}}{dt} = -x_{ij}^{L/R,+} + \left(1 - x_{ij}^{L/R,+}\right) I_{ij}^{L/R} - (1 + x_{ij}^{L/R,+}) \sum G_{pqij} I_{pq}^{L/R} , \tag{7}$$

$$\tau \frac{dx_{ij}^{L/R,-}}{dt} = -x_{ij}^{L/R,-} + \left(1 - x_{ij}^{L/R,-}\right) \sum G_{pqij} I_{pq}^{L/R} - (1 + x_{ij}^{L/R,-}) I_{ij}^{L/R} , \tag{8}$$

where constant $\tau = 0.1$, superscript L/R denotes that the cell belongs to the left or right monocular eye pathway, +/- denotes ON or OFF cell, indices i and j denote the position of the input on the retina image, $I_{ij}^{L/R}$ is the luminance of the left or right retinal image that represents the excitatory on-center, and G_{pqij} is a 5×5 Gaussian kernel that represents the inhibitory off-surround:

$$G_{pqij} = N \exp\left(-\frac{(p-i)^2 + (q-j)^2}{2} \right), \tag{9}$$

where N is a constant that normalizes the sum of the Gaussian kernel to 1. The firing threshold $x_{fire} = 0.02$ for fine spike detection.

V1 Monocular Simple Cells

Cells in V1 layer 4 are monocular simple cells that are sensitive to either dark-light or light-dark contrast polarity, but not both, depending on their receptive field structure. Their membrane potential is defined by

$$\begin{aligned} \frac{dr_{ijk}^{L/R,+}}{dt} = -\alpha r_{ijk}^{L/R,+} + \left(1 - r_{ijk}^{L/R,+}\right) \sum \mu (G_{pqijk}^{+} \hat{x}_{ij}^{L/R,+} + G_{pqijk}^{-} \hat{x}_{ij}^{L/R,-}) - \\ (1 + r_{ijk}^{L/R,+}) \sum \mu (G_{pqijk}^{+} \hat{x}_{ij}^{L/R,-} + G_{pqijk}^{-} \hat{x}_{ij}^{L/R,+}) \end{aligned} , \tag{10}$$

$$\begin{aligned} \frac{dr_{ijk}^{L/R,-}}{dt} = -\alpha r_{ijk}^{L/R,-} + \left(1 - r_{ijk}^{L/R,-}\right) \sum \mu (G_{pqijk}^{+} \hat{x}_{ij}^{L/R,-} + G_{pqijk}^{-} \hat{x}_{ij}^{L/R,+}) - \\ (1 + r_{ijk}^{L/R,-}) \sum \mu (G_{pqijk}^{+} \hat{x}_{ij}^{L/R,+} + G_{pqijk}^{-} \hat{x}_{ij}^{L/R,-}) \end{aligned} , \tag{11}$$

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, - indicates that the simple cell responds to light-dark contrast polarity, $k = 1, 2$ denotes horizontal or vertical orientations, $\alpha = 0.001$ denotes slow decay, conductance strength $\mu = 0.1$, $\hat{x}_{ij}$'s are spiking signals generated by LGN ON and OFF cells, and G's are 4×4 Gaussian kernels defined by

$$G_{ppqijk}^{+} = N^{+} \exp\left(-\frac{1}{2} \left(\frac{((p-i)\cos\theta + (q-j)\sin\theta + s)^2}{\sigma_s^2} + \frac{((p-i)\sin\theta - (q-j)\cos\theta)^2}{\sigma_l^2} \right) \right), \tag{12}$$

$$G_{ppqijk}^{-} = N^{-} \exp\left(-\frac{1}{2} \left(\frac{((p-i)\cos\theta + (q-j)\sin\theta - s)^2}{\sigma_s^2} + \frac{((p-i)\sin\theta - (q-j)\cos\theta)^2}{\sigma_l^2} \right) \right), \tag{13}$$

where N^{+} and N^{-} are constants that normalize the whole kernel to sum 1, $\theta = k\pi / 2$, $\sigma_s = 0.5$, $\sigma_l = 1$ and $s = \sigma_s / 2$.

These cells interact via on-center off-surround networks to sharpen V1 monocular boundaries and to inhibit nearby weak noisy signals:

$$\frac{ds_{ijk}^{L/R,+}}{dt} = -\alpha s_{ijk}^{L/R,+} + \left(1 - s_{ijk}^{L/R,+}\right)\eta \hat{r}_{ijk}^{L/R,+} - (1 + s_{ijk}^{L/R,+})\eta \sum W_{pqijk} (\hat{r}_{pqk}^{L/R,+} + \hat{r}_{pqk}^{L/R,-}), \qquad (14)$$

$$\frac{ds_{ijk}^{L/R,-}}{dt} = -\alpha s_{ijk}^{L/R,-} + \left(1 - s_{ijk}^{L/R,-}\right)\eta \hat{r}_{ijk}^{L/R,-} - (1 + s_{ijk}^{L/R,-})\eta \sum W_{pqijk} (\hat{r}_{pqk}^{L/R,+} + \hat{r}_{pqk}^{L/R,-}), \qquad (15)$$

where the variables s_{ijk} denote the membrane potentials of V1 layer 4 monocular cells, $\alpha = 0.001, \eta = 0.19,$ and the *W* terms are 3×3 kernels. W is defined by

$$W_{pqijk} = \begin{bmatrix} 0 & 0.5 & 0 \\ 0 & 0 & 0 \\ 0 & 0.5 & 0 \end{bmatrix}, \qquad (16)$$

for the horizontal orientation *k*. The kernel *W* for vertical orientation is obtained by appropriate rotation from horizontal orientation.

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, $b_{ijkd}^{+/-}$, of layer 3B simple cells obey the equation:

$$\frac{d}{dt} b_{ijkd}^{+/-}(t) = \mu_1 \left(\hat{s}_{(i+s)jk}^{L,+/-}(t) + \hat{s}_{(i-s)jk}^{R,+/-}(t)\right) - \alpha \left(\hat{q}_{ijkd}^{L,+/-}(t) + \hat{q}_{ijkd}^{R,+/-}(t)\right) \qquad (17)$$

where $\mu_1 = 0.3, \alpha = 0.49$ are excitatory and inhibitory conductance strength parameters, the $\hat{s}$ terms are spiking signals from layer 4 monocular simple cells, the $\hat{q}$ terms are spiking signals from 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.

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 farther 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 recurrent inhibitory input from other inhibitory interneuron that correspond to the same position, orientation and disparity. This enables the "two-against-one" computation in equation (17). The cell membrane potentials, $q_{ijkd}^{L/R,+/-}$, of the inhibitory interneurons obey:

$$\frac{d}{dt} q_{ijkd}^{L,+/-}(t) = \mu_2 \hat{s}_{(i+s)jk}^{L,+/-}(t) - \beta \hat{q}_{ijkd}^{L,-/+}(t-\varepsilon), \tag{18}$$

and

$$\frac{d}{dt} q_{ijkd}^{R,+/-}(t) = \mu_2 \hat{s}_{(i+s)jk}^{R,+/-}(t) - \beta \hat{q}_{ijkd}^{R,-/+}(t-\varepsilon), \tag{19}$$

where $\mu_2 = 0.3$, $\beta = 0.3$ are excitatory and inhibitory conductance strength parameters, and ε is a sufficiently small positive number that denotes there is a small signal delay between inhibitory interneurons. In the discrete-time simulation, $\varepsilon = \Delta t$, or one time step.

See Cao and Grossberg (2011) for a detailed analysis of how binocular simple cells obeying the matching equations (17-19) within proper parameter ranges can generate binocular matching properties with analog properties, including the ratio constraint on stereoscopic fusion (Smallman and McKee, 1995)

V1 Layer 2/3 Complex Cells

V1 layer 2/3 consists of both monocular and binocular complex cells. Binocular complex cells pool the spiking signals from binocular layer 3B simple cells of like orientation and both contrast polarities at each position. Their membrane potential, c_{ijkd}^{B}, thus obeys:

$$\frac{d}{dt} c_{ijkd}^{B} = -\alpha c_{ijkd}^{B} + (1 - c_{ijkd}^{B})(\widehat{b}_{ijkd}^{+} + \widehat{b}_{ijkd}^{-}), \tag{20}$$

where $\alpha = 0.001$, and $\widehat{b}$'s are spiking signals from layer 3B binocular simple cells.

Monocular complex cells pool the spiking signals from monocular layer 4 simple cells of like orientation and both contrast polarities at each position. Their membrane potential, $c_{ijk}^{L/R}$, thus obeys:

$$\frac{d}{dt} c_{ijk}^{L/R} = -\alpha c_{ijk}^{L/R} + (1 - c_{ijk}^{L/R})(\hat{s}_{ijk}^{L/R,+} + \hat{s}_{ijk}^{L/R,-}), \tag{21}$$

where $\alpha = 0.001$, and $\hat{s}$'s are spiking signals from layer 4 monocular simple cells.

V2 Layer 4 Cells

The left and right monocular inputs are combined with binocular input in layer 4 of V2 pale stripes. 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. 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 signals modulate the activities of V2 layer 4 boundary cells. The membrane potentials of V2 layer 4 cells, v_{ijkd}, thus obey:

$$\frac{d}{dt} v_{ijkd} = -\alpha v_{ijkd} + (1 - v_{ijkd}) \left(\gamma_b \hat{c}_{ijkd}^{B} + \gamma_m \left(\hat{c}_{(i+s)jk}^{L} + \hat{c}_{(i-s)jk}^{R} \right) \right) \left(1 + \mu \hat{f}_{ijkd} \right), \tag{22}$$

where $\alpha = 0.001$ is the decay rate, $\gamma_b = 2$ and $\gamma_m = 0.04$ represent the strength of the binocular and monocular connections, the $\hat{c}$ are binocular and monocular spiking signals from V1 layer

2/3 complex cell, s is the positional shift between left and right eye input that is defined in Table 1, $\hat{f}_{ijkd}$ is the spiking feedback signal from V2 left and right monocular surfaces, and $\mu = 5$ is the strength of the surface-to-boundary feedback. In order to generate graded properties, such as analog boundary strengths, all input spiking signals are of exponential decay type. In particular, V1 complex cell output signals $\hat{c}$ are first represented by (6) with $\alpha = 500$, then input into V2 layer 4 cells.

V2 Layer 2/3 Complex Cells

The V2 layer 2/3 cells receive input from V2 layer 4. 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, thereby computing "two-against-one" properties during boundary grouping. The rate-based models have shown that this balance of excitation and inhibition helps to implement a bipole property that controls boundary grouping (Grossberg and Mingolla, 1985a, 1985b; Grossberg, 1999a; Grossberg and Raizada, 2000). The disparity filter that solves the correspondence problem is realized as part of the inhibitory interactions that control perceptual grouping. In order to prevent the grouping process from driving cell firing too much, a spike-dependent refractory period (term R in equations (23) and (29)) limits the maximum firing rate of these spiking neurons. 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* accordingly obeys the equation:

$$\begin{aligned}\frac{d}{dt} g_{ijkd} &= -\alpha g_{ijkd} + (1 - g_{ijkd})\Big(\beta_1 \hat{v}_{ijkd} + \beta_2 (\hat{H}^1_{ijkd} + \hat{H}^2_{ijkd})\Big) \\ &\quad -(1 + g_{ijkd})\Big(\gamma_1 (\hat{Q}^1_{ijkd} + \hat{Q}^2_{ijkd}) + \gamma_2 \hat{R}_{ijkd} + \gamma_3 \hat{P}_{ijkd}\Big)\end{aligned} \tag{23}$$

where $\alpha = 0.001$, $\beta_1 = 0.3$, $\beta_2 = 0.21$, $\gamma_1 = \beta_2 / 2$, $\gamma_2 = 0.3$, $\gamma_3 = 0.1$, and $\hat{v}_{ijkd}$ are the spiking inputs from V2 layer 4 cells. All spiking signals that input into the bipole cell are exponential decay type.

The V2 layer 2/3 collinear bipole cells receive long-range input from other collinear and coaxial bipole cells at nearby positions with the same disparity preference. Term $\hat{H}^m_{ijkd}$ ($m = 1, 2$) is the spiking input from branch *m* of the bipole cell at position *(i,j)*, orientation *k* and disparity *d*:

$$\hat{H}^m_{ijkd} = \sum_{pq} W^m_{pqijk} \hat{g}_{pqkd}, \quad m = 1, 2, \tag{24}$$

where W^m_{pqijk} is long-range connection weight, defined by an 11×11 elongated Gaussian kernel as follows

$$W^1_{pqij1} = N^1 \left[sign(i-p) \exp\left(-\left(\frac{(i-p)^2}{\sigma_p^2} + \frac{(j-q)^2}{\sigma_q^2} \right) \right) \right]^+, \tag{25}$$

$$W^2_{pqij1} = N^2 \left[sign(p-i) \exp\left(-\left(\frac{(i-p)^2}{\sigma_p^2} + \frac{(j-q)^2}{\sigma_q^2} \right) \right) \right]^+, \tag{26}$$

for horizontal orientation. Here N^m ($m = 1, 2$) is a normalization constant such that the sum of all elements in kernel W^m_{pqijk} equals 1, $sign(x) = 1$ if $x > 0$, -1 if $x < 0$, and 0 otherwise,

$\sigma_p = 20$, and $\sigma_q = 0.3$. The kernel for vertical orientation is obtained by appropriate rotation. $\hat{g}_{pqkd}$ in (24) is the exponential decay type spiking signal generated by the bipole cell at position (*p*,*q*) with orientation *k* and disparity *d*.

Terms $\hat{Q}^m_{ijkd}$ ($m = 1,2$) are the inhibitory inputs from the di-synaptic inhibitory interneurons that control the bipole property of perceptual grouping, defined by:

$$\frac{d}{dt}Q^1_{ijkd} = \beta_q \hat{H}^1_{ijkd} - \gamma_q \hat{Q}^2_{ijkd}, \tag{27}$$

$$\frac{d}{dt}Q^2_{ijkd} = \beta_q \hat{H}^2_{ijkd} - \gamma_q \hat{Q}^1_{ijkd}, \tag{28}$$

where $\beta_q = 0.21$, $\gamma_q = 0.21$, and $\hat{H}$'s are defined in (24). Note that the firing threshold for the inhibitory cell Q^m_{ijkd} is 0.1 so that the inhibitory cells fire faster than the targeted bipole cell. This ensures that the bipole property that controls boundary grouping is satisfied. Note also the similarity of equations (27) and (38) with the recurrent inhibitory interneurons in equations (18) and (19) that realize the disparity-selective responses of binocular simple cells.

Term $\hat{R}_{ijkd}$ is a spike-dependent refractory term which prevents a cell from firing at too high a frequency:

$$\frac{d}{dt}R_{ijkd} = -\alpha_R R_{ijkd} + \beta_R \hat{g}_{ijkd}, \tag{29}$$

where $\alpha_R = 50$ denotes a fast decay rate, $\beta_R = 0.11$, and $\hat{g}$ defines spikes that are generated by the bipole cell. By (29), the refractory term decays at a fast exponential rate after each spike boosts its value.

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	-

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 $\hat{P}_{ijkd}$ in (23) is the inhibition across disparities along the lines-of-sight:

$$\hat{P}_{ijkd} = \sum_{d' \neq d} M_{dd'}(\hat{g}_{(i+s'-s)jkd'} + \hat{g}_{(i+s-s')jkd'}), \tag{30}$$

where $M_{dd'}$ 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 $\hat{g}_{(i+s'-s)jkd'}$ and $\hat{g}_{(i+s-s')jkd'}$ 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). The disparity filter $\hat{P}_{ijkd}$ 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).

V2 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 control monocular surface filling-in. The boundary signals received from V2 layer 2/3 create resistive barriers to the filling-in process. Similar to the rate-based model (e.g., Cao and Grossberg, 2005; Cohen and Grossberg, 1984; Fang and Grossberg, 2009; Grossberg, 1984, 1994, 1997; Grossberg and Howe, 2003; Grossberg and Kelly, 1999; Grossberg and Todorovic, 1988), filling-in is modeled by a diffusion equation. The membrane potentials of an ON surface cell $F_{ijd}^{L/R,+}$ and an OFF surface cell $F_{ijd}^{L/R,-}$ are defined by

$$\varepsilon \frac{d}{dt} F_{ijd}^{L/R,+} = -\alpha F_{ijd}^{L/R,+} + \eta \sum_{(p,q)\in N_{ij}} \left(\hat{F}_{pqd}^{L/R,+} - \hat{F}_{ijd}^{L/R,+}\right)\Psi_{pqijd} + \mu \hat{x}_{ijd}^{L/R,+}, \tag{31}$$

$$\varepsilon \frac{d}{dt} F_{ijd}^{L/R,-} = -\alpha F_{ijd}^{L/R,-} + \eta \sum_{(p,q)\in N_{ij}} \left(\hat{F}_{pqd}^{L/R,-} - \hat{F}_{ijd}^{L/R,-}\right)\Psi_{pqijd} + \mu \hat{x}_{ijd}^{L/R,-}, \tag{32}$$

respectively, where rate parameter $\varepsilon << 1$, implies that the monocular surface filling-in process in the thin stripes is faster than the grouping process in V2 layer 2/3. In the simulation, for every iteration in the grouping process, 60 iterations were done in the filling-in process. The decay rate $\alpha = 2000$ denotes a fast decay, parameters $\eta = 0.5$ and $\mu = 0.25$ are the strengths of diffusion and bottom-up input, $\hat{F}_{ijd}^{L/R,+}$ and $\hat{F}_{ijd}^{L/R,-}$ are the firing signals of the left/right ON and OFF surface cells at position (*i,j*) and disparity *d*; and N_{ij} is the set of the nearest-neighbor locations of *(i,j)*:

$$N_{ij} = \{(i, j-1), (i-1, j), (i+1, j), (i, j+1)\}; \tag{33}$$

and $\hat{x}_{ijd}^{L/R,+}$ and $\hat{x}_{ijd}^{L/R,-}$ are the spiking signals from left/right LGN, given by:

$$\hat{x}_{ijd}^{L,+/-} = \hat{x}_{(i+s)j}^{L,+/-}, \tag{34}$$

and

$$\hat{x}_{ijd}^{R,+/-} = \hat{x}_{(i-s)j}^{R,+/-}, \tag{35}$$

where *s* is the positional shift defined in Table 1.

The boundary-gating coefficients, Ψ_{pqijd}, in (31) and (32) respond to the boundary signals from V2 layer 2/3 by creating barriers to filling-in. These barriers inhibit the diffusion at the positions of the boundaries. A challenge for surface filling-in by spiking neurons is how to define the diffusion-gating coefficient, Ψ_{pqijd}. It cannot be defined as a steady-state inverse of boundaries, as in the rate-based model (Grossberg and Todorović, 1988), because the boundary signals from V2 layer 2/3 are discrete spikes. The boundary-gating process, Ψ_{pqijd}, is instead defined by a gating signal Ψ_{pqijd} that continuously inactivates, or habituates (Abbott et al., 1997; Francis et al., 1994; Grossberg, 1968, 1969. 1980), in response to spikes $\hat{G}_{pqijd}$:

$$\frac{d}{dt}\Psi_{pqijd} = \rho(1-\Psi_{pqijd}) - \beta\hat{G}_{pqijd}\Psi_{pqijd}\,, \tag{36}$$

where $\rho = 1$ denotes how fast Ψ_{pqijd} can grow to its maximum value 1 when there is no boundary spike from V2 layer 2/3, $\beta = 1000$ denotes how fast Ψ_{pqijd} will decrease (inactivate) to zero when a boundary spike arrives, and $\hat{G}_{pqijd}$ is the sum of nearby boundary signals defined by

$$\hat{G}_{pqijd} = \begin{cases} \hat{g}_{(i-0.5)(j-0.5)d} + \hat{g}_{(i-0.5)(j+0.5)d}\,, & if\ \ p = i-1, q = j, \\ \hat{g}_{(i+0.5)(j-0.5)d} + \hat{g}_{(i+0.5)(j+0.5)d}\,, & if\ \ p = i+1, q = j, \\ \hat{g}_{(i-0.5)(j-0.5)d} + \hat{g}_{(i+0.5)(j-0.5)d}\,, & if\ \ p = i, q = j-1, \\ \hat{g}_{(i-0.5)(j+0.5)d} + \hat{g}_{(i+0.5)(j+0.5)d}\,, & if\ \ p = i, q = j+1, \end{cases} \tag{37}$$

where the boundary terms $\hat{g}$'s sum over all orientations of bipole cell output spiking signals:

$$\hat{g}_{ijd} = \sum_k \hat{g}_{ijkd}\ , \tag{38}$$

Note that (36) describes the dynamics of the diffusion-gating coefficients, Ψ_{pqijd} . It is not an integrate-and-fire equation, and hence there is no spike generated by (36).

Equation (36) shows that any large spiking boundary signal at the nearest neighbor locations reduces the diffusion coefficient Ψ_{pqijd} and thereby blocks filling-in.

Surface-to-Boundary Feedback Signals

The V2 monocular surfaces in the thin stripes generate surface contour signals that generate surface-to-boundary feedback signals to layer 4 cells in the V2 pale stripes; see equation (22). Output signals from the filled-in activities in the V2 thin stripes are derived from oriented filters that define the surface contour signals:

$$\begin{aligned} \frac{df_{ijkd}^{L/R,+}}{dt} = -\alpha f_{ijkd}^{L/R,+} + \left(1 - f_{ijkd}^{L/R,+}\right)\sum \mu(G_{pqij}^{+}\hat{F}_{ijd}^{L/R,+} + G_{pqij}^{-}\hat{F}_{ijd}^{L/R,-}) - \\ (1 + f_{ijk}^{L/R,+})\sum \mu(G_{pqij}^{+}\hat{F}_{ijd}^{L/R,-} + G_{pqij}^{-}\hat{F}_{ijd}^{L/R,+}) \end{aligned} \tag{39}$$

and

$$\begin{aligned} \frac{df_{ijkd}^{L/R,-}}{dt} = -\alpha f_{ijkd}^{L/R,-} + \left(1 - f_{ijkd}^{L/R,-}\right)\sum \mu(G_{pqij}^{+}\hat{F}_{ijd}^{L/R,-} + G_{pqij}^{-}\hat{F}_{ijd}^{L/R,+}) - \\ (1 + f_{ijkd}^{L/R,-})\sum \mu(G_{pqij}^{+}\hat{F}_{ijd}^{L/R,+} + G_{pqij}^{-}\hat{F}_{ijd}^{L/R,-}), \end{aligned} \tag{40}$$

where $\alpha = 0.001$, $\mu = 0.1$, $\hat{F}$'s are the spiking signals from V2 monocular surface cells, and kernels G_{pqij}^{+} and G_{pqij}^{-} are defined by (12) and (13). Note that the firing rate in (39) and (40) is 0.1 for fine spiking detection.

The surface-to-boundary spiking signals $\hat{f}_{ijkd}$ are then defined by:

$$\hat{f}_{ijkd} = \hat{f}^{L,+}_{ijkd} + \hat{f}^{L,-}_{ijkd} + \hat{f}^{R,+}_{ijkd} + \hat{f}^{R,-}_{ijkd} , \tag{41}$$

where $\hat{f}^{L/R,+/-}_{ijkd}$ are spiking signals from (39) and (40). Due to the discrete property of spikes, the surface contour feedback signals $\hat{f}_{ijkd}$ are defined by exponential decay type spikes before feeding back to V2 layer 4 boundary cells in (22). The exponential decay function is defined slightly differently from (6) in which its maximum value is 1:

$$e(t) = e^{-\alpha(t-t_f)} , \tag{42}$$

where $\alpha = 10$ denotes a slow decay and t_f is the last firing time as in (6).

V4 surface

V4 surface cells receive lightness signals from the LGN via V1 blobs and V2 thin stripes, and boundary signals from V2 layer 2/3 via V1 interblobs and V2 pale stripes. Each such cell combines the monocular lightness signals from the two eyes that correspond to the same 3D location. Its binocular ON and OFF lightness inputs, $\hat{z}^{+/-}_{ijd}$, are given by:

$$\hat{z}^{+/-}_{ijd} = \hat{x}^{L,+/-}_{ijd} + \hat{x}^{R,+/-}_{ijd} , \tag{43}$$

where $\hat{x}^{L,+/-}_{ijd}$ and $\hat{x}^{R,+/-}_{ijd}$ are defined in (34) and (35). The membrane potentials of V4 ON and OFF cells, $w^{+/-}_{ijd}$, are then modeled by an equation similar to (31) and (32):

$$\varepsilon \frac{d}{dt} w^{+/-}_{ijd} = -\alpha w^{+/-}_{ijd} + \eta \sum_{(p,q)\in N_{ij}} \left(\hat{w}^{+/-}_{pqd} - \hat{w}^{+/-}_{ijd}\right)\Psi_{pqijd} + \mu \hat{z}^{+/-}_{ijd} , \tag{44}$$

where all parameters and Ψ_{pqijd} are the same as in (31) and (32).

The final 3D surface percept, w_{ijd}, in the simulation is computed by the time-average of the difference of firing rates of ON and OFF cells:

$$w_{ijd} = \int_{t_0}^{T} \left(\hat{w}^{+}_{ijd}(t) - \hat{w}^{-}_{ijd}(t)\right) dt , \tag{45}$$

where t_0 and T are the start and end time. In the simulation, $t_0 = 500$ and $T = 2000$. In other words, 2000 iterations were run for each experiment, and spikes were counted from the 500th iteration when the network approaches its stable state.

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

Cao, Y. and Grossberg, S. (2011). Laminar cortical computation of stereo disparity by spiking neurons. Submitted for publication.

Cao, Y., Grossberg, S., and Markowitz, J. (2011). How does the brain rapidly learn and reorganize view-invariant and position-invariant object representations in the inferotemporal cortex? *Neural Networks*, *24*, 1050-1061.

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. (2009) From stereogram to surface: How the brain sees the world in depth. *Spatial Vision*, 22, 45-82

Fazl, A., Grossberg, S., and Mingolla, E. (2009). View-invariant object category learning, recognition, and search: How spatial and object attention are coordinated using surface-based attentional shrouds. *Cognitive Psychology*, 58, 1-48.

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. (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. (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. (2009). Cortical and subcortical predictive dynamics and learning during perception, cognition, emotion, and action. *Philosophical Transactions of the Royal Society of London*, 364, 1223-1234.

Grossberg, S. and Hong, S. (2006). A neural model of surface perception: Lightness, anchoring, and filling-in. *Spatial Vision,* 19, 263-321.

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., 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 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 Versace, M. (2008). Spikes, synchrony, and attentive learning by laminar thalamocortical circuits. *Brain Research*, 1218, 278-312.

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. (2005). Laminar Cortical Dynamics of 3D Surface Perception: Stratification, Transparency, and Neon Color Spreading. *Vision Research,45,* 1275-1743.

Grossberg, S., Yazdanbakhsh, A., Cao, Y., and Swaminathan, G. (2008). How does binocular rivalry emerge from cortical mechanisms of 3-D vision? *Vision Research*, 48, 2232-2250.

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 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.

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.

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

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.

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.

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