FROM STEREOGRAM TO SURFACE:

How the Brain Sees the World in Depth

Liang Fang and Stephen Grossberg
Department of Cognitive and Neural Systems
Center for Adaptive Systems
and
Center of Excellence for Learning in Education, Science, and Technology
Boston University
677 Beacon St., Boston, MA 02215
Phone: 617-353-7858 or -7857
Fax: 617-353-7755
E-mail: lfang@cns.bu.edu, steve@bu.edu
To appear in Spatial Vision, Special issue on Unresolved Questions in Stereopsis

Corresponding author: Stephen Grossberg

Submitted: April, 2007
Revised: August, 2007

Acknowledgements
L.F. was supported in part by the Air Force Office of Scientific Research (AFOSR F49620-01-1-0397), the National Science Foundation (NSF EIA-01-30851 and NSF SBE-0354378), and the Office of Naval Research (ONR N00014-01-1-0624). S.G. was supported in part by the National Science Foundation (NSF SBE-0354378) and the Office of Naval Research (ONR N00014-01-1-0624).

ABSTRACT

How do we consciously see surfaces infused with lightness and color at the correct depths? Random Dot Stereograms (RDS) probe how binocular disparity between the two eyes can generate conscious surface percepts. Dense RDS do so despite the fact that they include multiple false binocular matches. Sparse stereograms do so across large contrast-free regions with no binocular matches. Stereograms that define occluding and occluded surfaces lead to surface percepts wherein partially occluded textured surfaces are completed behind occluding textured surfaces at a spatial scale much larger than that of the texture elements themselves. Earlier models suggest how the brain detects binocular disparity, but not how RDS generate conscious percepts of 3D surfaces. This article proposes a neural model that predicts and simulates how the layered circuits of visual cortex generate 3D surface percepts using interactions between boundary and surface representations that obey complementary computational rules. The model clarifies how interactions between layers 4, 3B, and 2/3A in V1 and V2 contribute to stereopsis, and proposes how 3D perceptual grouping laws in V2 interact with 3D surface filling-in operations in V1, V2, and V4 to generate 3D surface percepts in which figures are separated from their backgrounds.

Keywords: stereopsis, visual cortex, surface perception, figure-ground separation, LAMINART model.

Introduction

When we view a 3D scene, we effortlessly perceive the world in depth. The positional differences of an object's projections on an observer's left and right retinas, or their *binocular disparity*, is one cue for perceiving depth (Helmholtz, 1866; Howard and Rogers, 2002). *Random-Dot-Stereograms* (RDS) have been used to study stereopsis and 3D vision since Julesz introduced them to show how binocular disparity can cause 3D percepts without monocular cues (Julesz, 1971). How does the visual cortex generate 3D surface percepts from RDS, and what are the perceptual units that are used to do so?

Most previous stereopsis models (Dev, 1975; Fleet, Wagner, and Heeger, 1996; Grimson, 1981; Julesz, 1971; Marr and Poggio, 1976; Ohzawa, DeAngelis, and Freeman, 1990; Oian, 1994; Sperling, 1970) analysed how left and right eye contours are matched (the *correspondence problem*) (Howard and Rogers, 2002; Julesz, 1971), but not how matching leads to 3D perceptual groupings that support conscious percepts of surface lightness and color. Here, three types of 3D surface percepts that are generated by RDS are explained and simulated in order to illuminate crucial brain processes of 3D visual perception. A detailed neural model, called 3D LAMINART, predicts how the Lateral Geniculate Nucleus (LGN) and the laminar circuits of visual cortical areas V1, V2, and V4 interact to generate such percepts by forming 3D boundaries and surfaces in the interblob and blob cortical streams, respectively. These boundaries and surfaces obey complementary computational rules, and are not the independent modules that many previous models espouse. The present study extends the explanatory and predictive range of the 3D LAMINART model, as summarized in the Discussion section, to include the following types of percepts:

Dense RDS. Dense RDS contain crowded features that make the correspondence problem hard to solve by including many false binocular matches (Howard and Rogers, 2002; Julesz, 1971). The 3D LAMINART model separates objects and their surface lightnesses in depth in response to a dense stereogram. Figure 1 illustrates such a stereogram and the model's ability to generate a 3D surface percept from it. How the model does this is explained below.

Sparse RDS. Sparse RDS contain widely separated features that generate percepts of continuous surfaces in depth across large feature-free image regions, whose depth is locally ambiguous. Local filtering of contrast features can compute binocular disparities only at the matched edges of the sparse image features. The model proposes how a 3D perceptual grouping process responds to the filtered sparse features to form connected boundaries at multiple depths. These boundaries induce and contain filling-in of surface lightness at different depths by a process of 3D surface capture. Figure 2 illustrates such a stereogram and the model's simulated percept.

Dense RDS that induce emergent occlusion. Dense RDS that implicitly define a partially occluded object illustrate how multiple-scale boundary processing occurs: small-scale boundary groupings capture consciously seen small-scale texture lightnesses, while large-scale groupings form behind the occluder and perceptually link the visible parts of the occluded object. These completed boundaries are amodally recognized, but they are not seen with visible surface lightness. Figure 3 illustrates such a stereogram and the model's simulated percept.

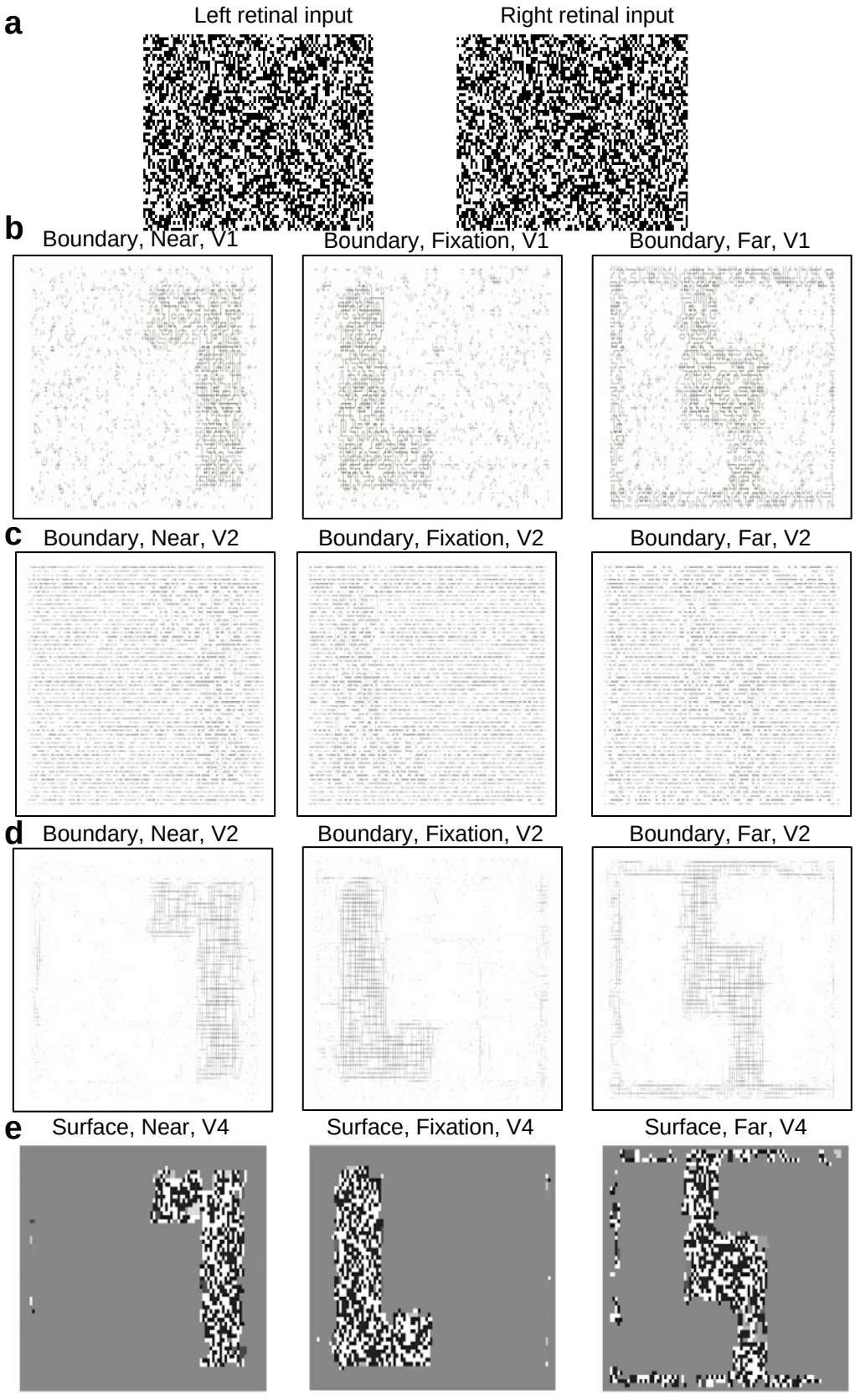

a
Left retinal input
Right retinal input
b
Boundary, Near, V1
Boundary, Fixation, V1
Boundary, Far, V1
c
Boundary, Near, V2
Boundary, Fixation, V2
Boundary, Far, V2
d
Boundary, Near, V2
Boundary, Fixation, V2
Boundary, Far, V2
e
Surface, Near, V4
Surface, Fixation, V4
Surface, Far, V4

Figure 1. Dense random-dot-stereogram: (a) Retinal inputs. (b) Boundaries at model V1 binocular complex cells. (c) Boundaries at model V1 binocular complex cells when blob-to-interblob feedback is prevented. (d) Boundaries at layer 2/3 binocular bipole cells within model V2 pale stripes. (e) Visible surfaces at model V4 cells. See text for details. When two images are successfully fused, a white square dotted with black is perceived to be floating over a white background that is also dotted with black. The square is perceived to be slightly whiter than the background. The question is why the feature-absent spaces between the dots are perceived in the correct depths. (b)

Methods

What are the units of visual perception? How does the brain generate visible 3D surface percepts from natural scenes and from stereograms? The 3D LAMINART model builds on the prediction (Cohen and Grossberg, 1984; Grossberg 1987; Grossberg and Mingolla, 1985; Grossberg and Todorović, 1988), which many subsequent experiments have supported, that *boundaries* and *surfaces* are the brain's perceptual units. This prediction represented a radical break in vision theory because it contradicted the popular hypothesis that the brain sees by using *independent modules*. The concept of independent modules tried to explain how visual properties (e.g., form, color, motion, depth) use specialized processes to be computed. The brain's organization into distinct anatomical areas and processing streams supports the idea that brain processing is specialized (DeYoe and van Essen, 1988). However, specialization does not imply independence. Independent modules should be able to fully compute a property like form, color, motion, or depth on its own. It is known, in contrast, that perceptual qualities interact, including form and motion, and brightness and depth (Egusa, 1983; Faubert and von Grunau, 1995; Gilchrist, 1977; Kanizsa, 1974; Pessoa, Beck, and Mingolla, 1996; Smallman and McKee, 1995). The hypothesis of independent modules is not supported at the anatomical level either, since different cortical processing streams are known to interact with one another.

Complementary boundary and surface streams. The prediction that the brain computes boundaries and surfaces embodies a different view of brain specialization, which helps to explain why cortical processing streams interact; namely, specialization is achieved through *complementary computing* (Grossberg, 2000). The complementary properties of boundaries and surfaces (Figure 4) that support visual perception are predicted to be carried out by the LGN-(V1 interblob)-(V2 interstripe)-V4 boundary stream, and the LGN-(V1 blob)-(V2 thin stripe)-V4 surface stream, respectively. The ability of cells in one stream to compute one sort of property (e.g., a perceptual boundary) prevents it, in principle, from computing a complementary property (e.g., a perceptual surface). The streams interact at multiple stages of processing to overcome their complementary weaknesses. A neurophysiologist who records a cell in one stream would therefore have considerable difficulty separating properties that are computed directly in that stream from properties that are projected to the stream from cells in the complementary stream. Our model (Figures 5 and 6) predicts how such inter-stream interactions in both V1 (between blobs and interblobs) and V2 (between thin stripes and pale stripes) enable conscious percepts of 3D surfaces to be seen in response to RDS and other visual stimuli.

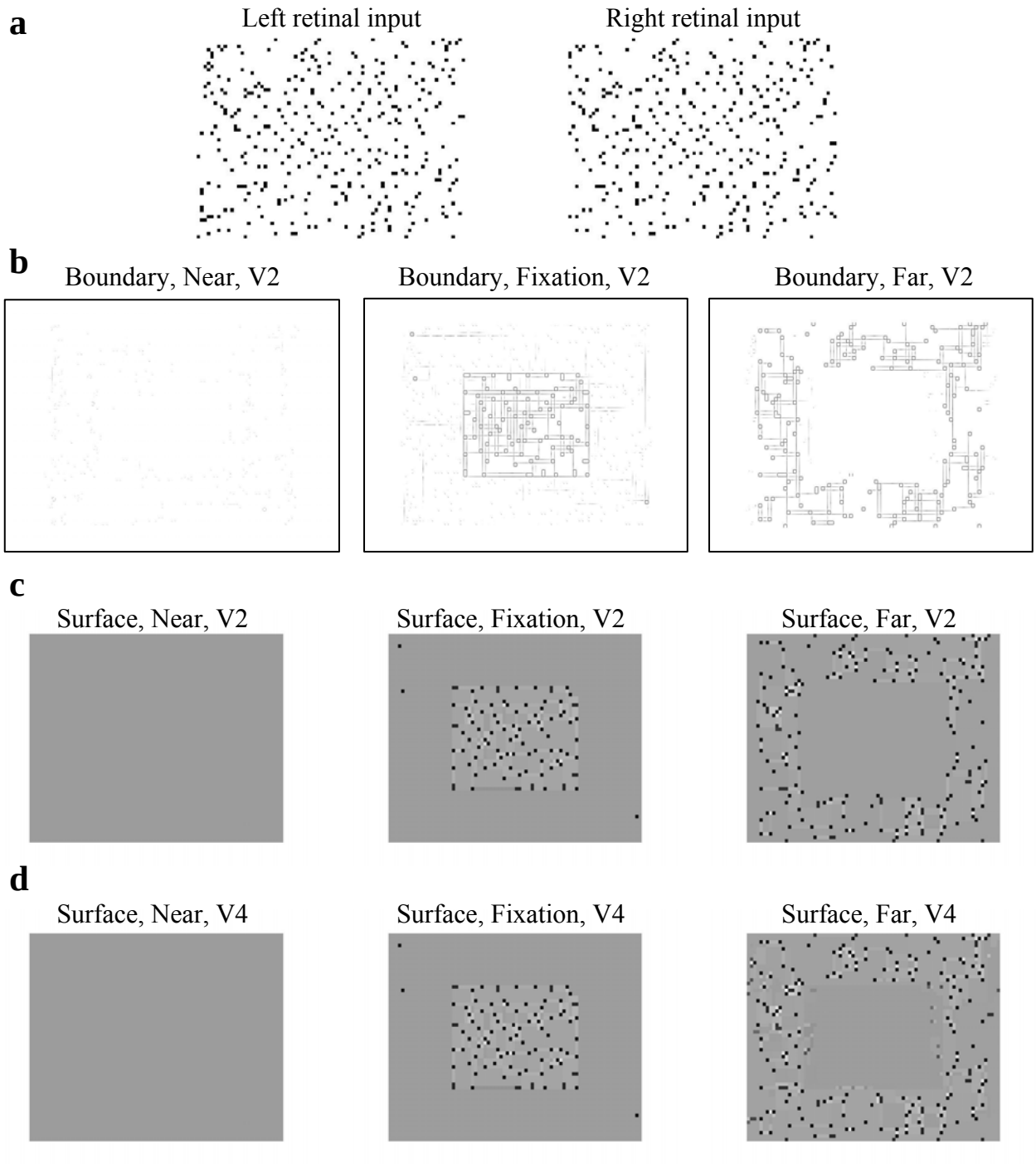


Figure 2. Sparse Random-dot-stereogram: (a) Retinal inputs. (b) Boundaries at layer 2/3 bipole cells in model V2 pale stripes. (c) Amodal surfaces at V2 thin stripes. (d) Visible surfaces at V4 cells. See text for details.

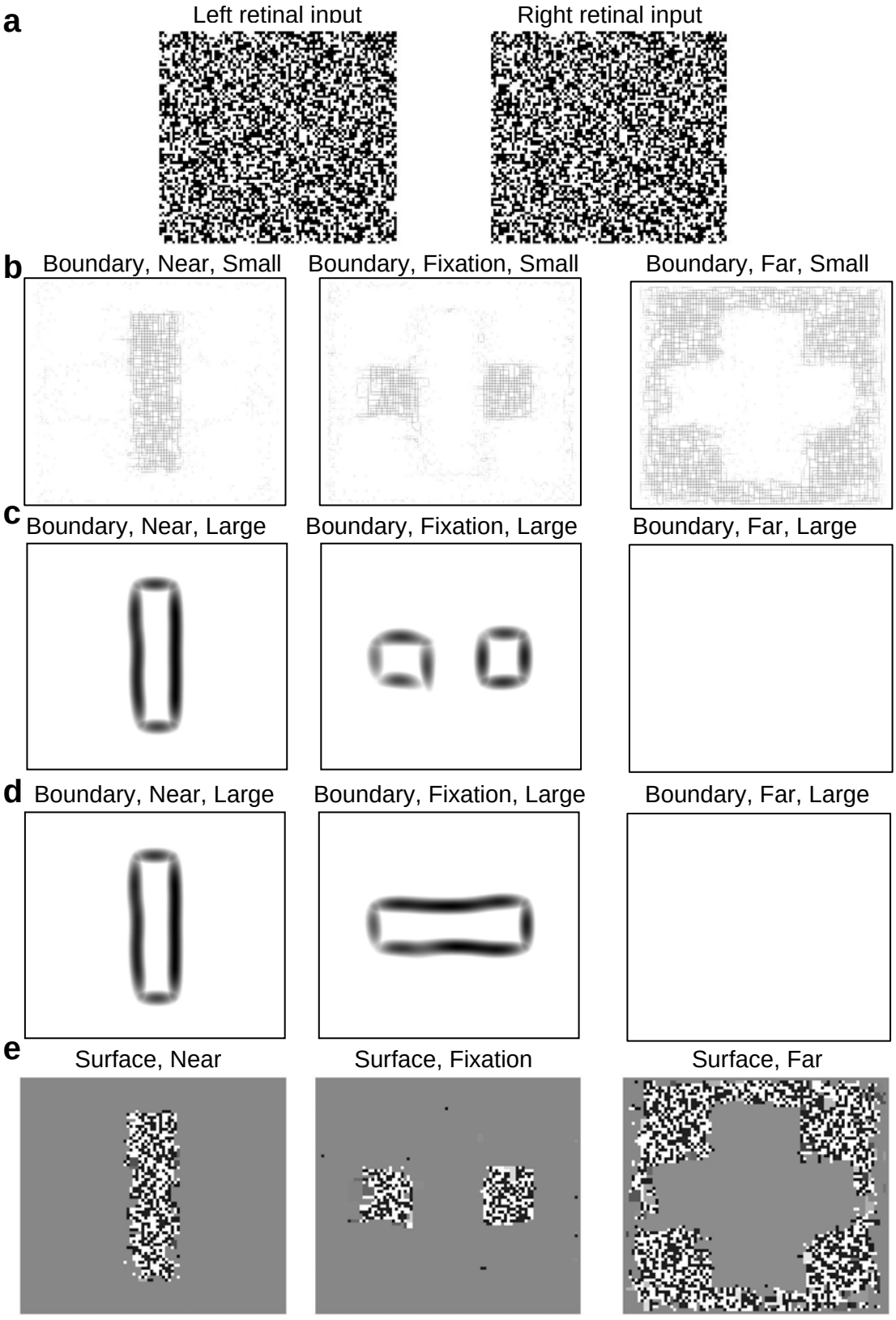
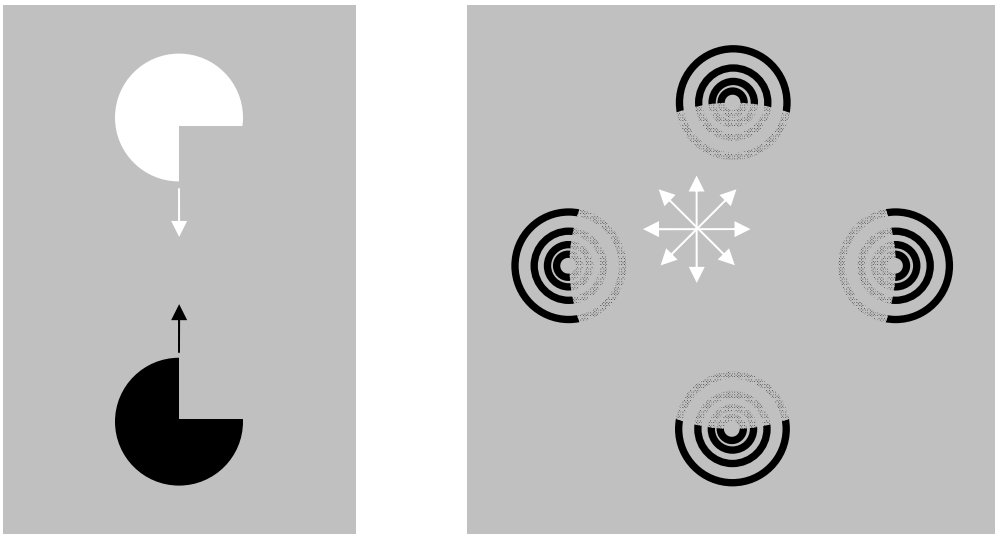


Figure 3. Random-dot-stereogram with partial occlusion. This percepts illustrates the figure-ground percept when a dense stereogram induces a textured surface that

partially occludes another textured surface: (a) Retinal inputs: When two images are successfully fused, a textured vertical bar is perceived to be nearer than a textured horizontal bar and to occlude the central part of the horizontal bar, with both bars nearer than the textured background. The visible parts of the textured horizontal bar on the flanks of the vertical bar are perceived to be perceptually grouped together to give the percept of a partially occluded bar, instead of two separated squares. (b) Small-scale boundaries in bipole cells of layer 2/3 model V2 pale stripes. (c) Large-scale boundaries in bipole cells of layer 2/3 model V2 pale stripes. Simulation before surface-to-boundary feedback occurs from V2 thin stripes to pale stripes. (d) Large-scale boundaries after surface-to-boundary feedback occurs. (e) Visible surfaces in depth at model cells in V4. See text for details.



Boundary: *Completion*

Oriented
Formed inwardly
Insensitive to contrast-polarity

Surface: *Filling-in*

Un-oriented
Formed outwardly
Sensitive to contrast-polarity

Figure 4. Boundaries and surfaces obey complementary computational rules. Boundary completion (left) is sensitive to orientation, occurs inwardly between two or more approximately collinear and like-oriented inducers, and pools over opposite contrast polarities, rendering it insensitive to contrast polarity. Surface filling-in (right) spreads in an unoriented way, outwardly from individual inducers, and is sensitive to contrast polarity, hence can support a conscious percept of visible surface features.

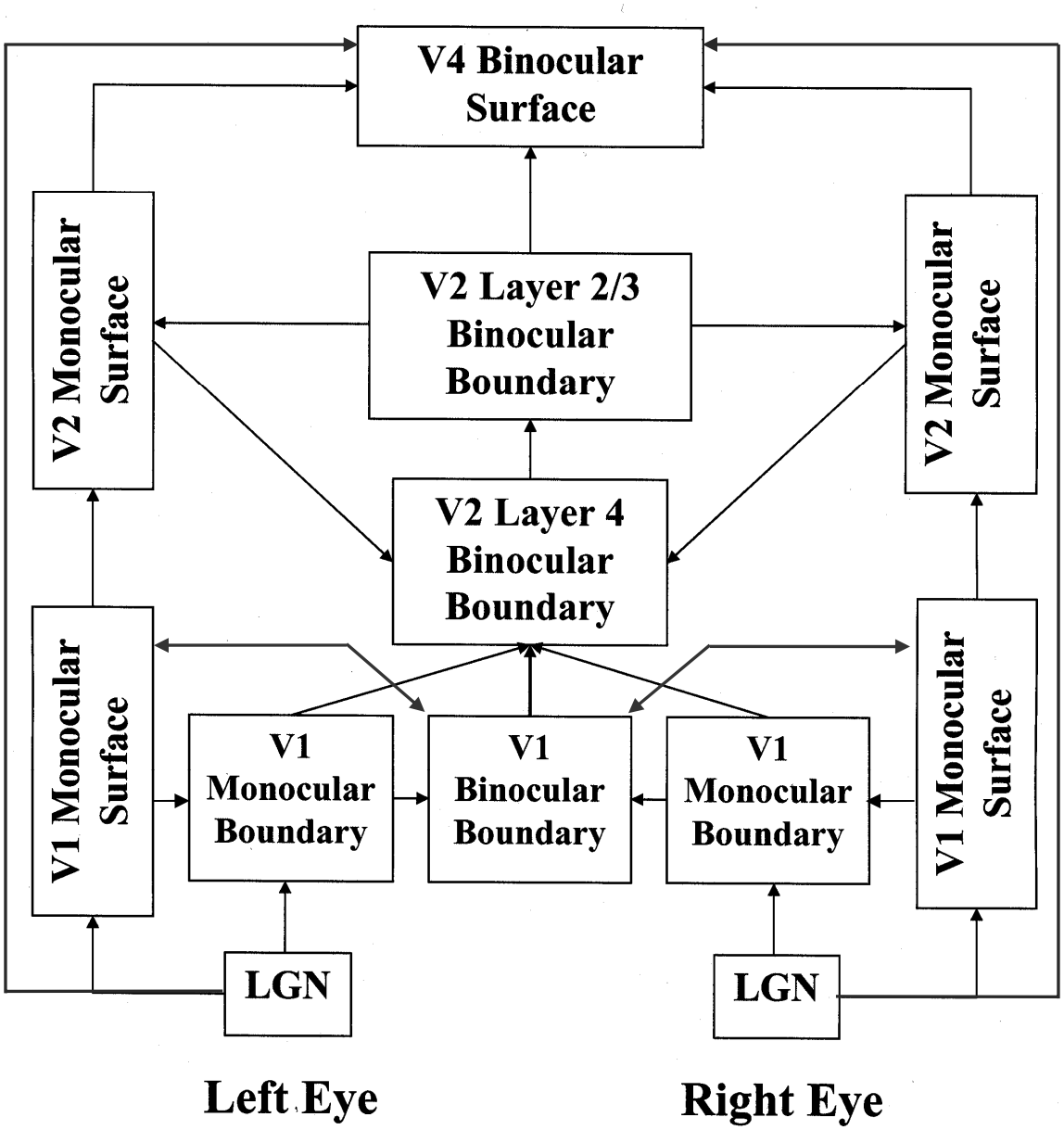


Figure 5. 3D LAMINART model macrocircuit. Feedback connections between boundaries and surfaces overcome the complementary deficiencies of each stream and lead to consistent conscious percepts. See text for details.

In order for the correct percepts to be seen, the brain needs to interactively solve a *boundary correspondence problem* (BCP) and a *surface correspondence problem* (SCP). A novel model prediction about the BCP is that elimination of false binocular boundary matches depends on 3D perceptual grouping circuits in layer 2/3 of V2 pale stripes. A novel model prediction about the SCP is that elimination of depth-ambiguous horizontal boundaries depends upon V1 surface-to-boundary feedback from blobs to interblobs (Figure 6), as well as the V2 surface-to-boundary feedback from thin stripes to pale stripes that was earlier used to explain other data.

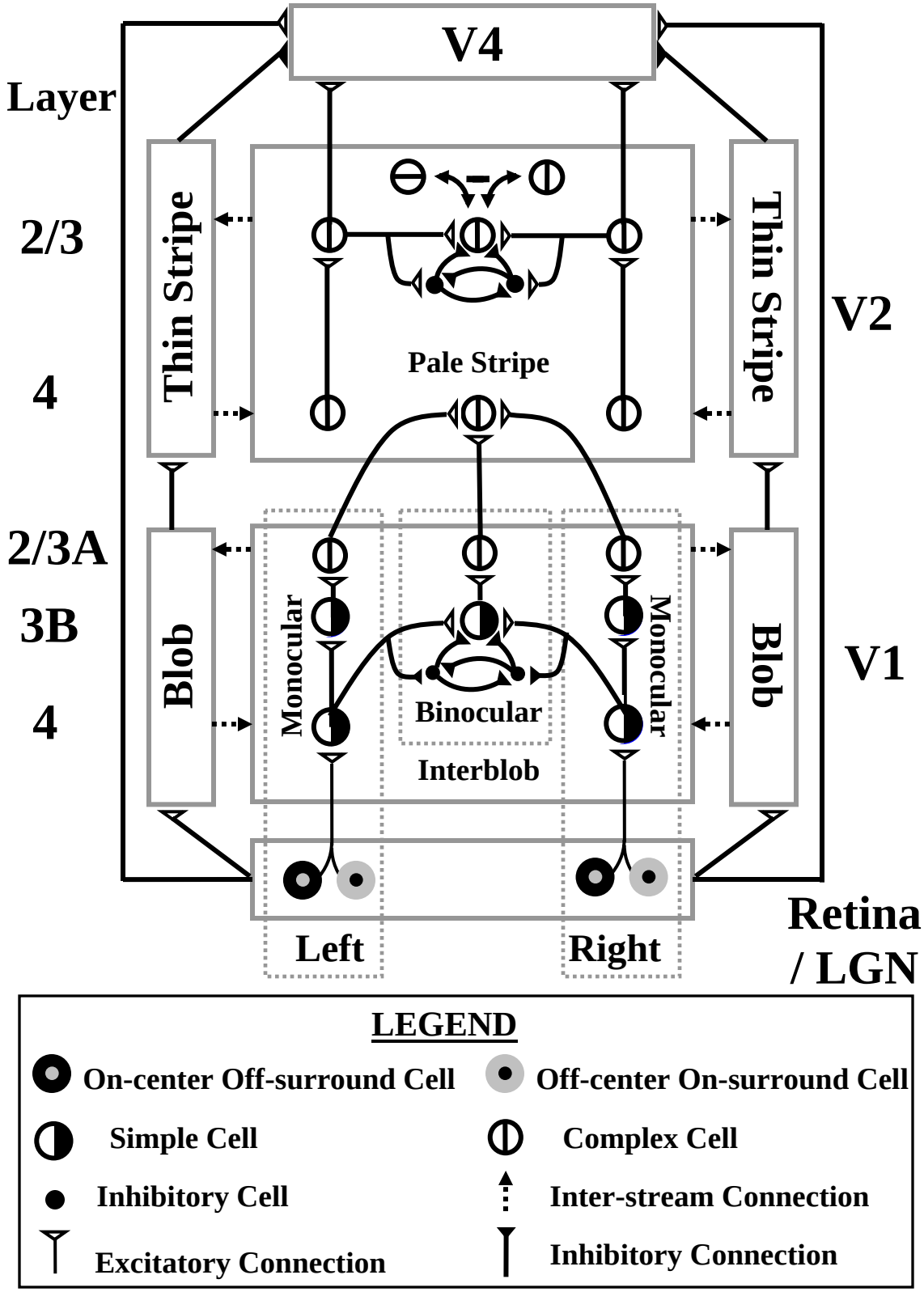

V4
Layer
2/3
4
2/3A
3B
4
Thin Stripe
Thin Stripe
Pale Stripe
V2
Blob
Blob
Monocular
Monocular
Binocular
Interblob
V1
Left
Right
Retina
/ LGN
LEGEND
On-center Off-surround Cell
Off-center On-surround Cell
Simple Cell
Complex Cell
Inhibitory Cell
Inter-stream Connection
Excitatory Connection
Inhibitory Connection

Figure 6. 3D LAMINART model circuit diagram. The model consists of a boundary stream that includes V1 interblobs, V2 pale stripes, and part of V4, and computes 3D perceptual groupings in different scales; and a surface stream that includes V1 blobs, V2 thin stripes, and part of V4, and computes 3D surfaces that are infused with lightness in depth. Both the boundary and surface streams receive illuminant-discounted signals from LGN cells with center-surround receptive fields, and both converge in V4, where visible 3D surfaces are consciously seen that are separated from their backgrounds. Model V2 and V4 also output to inferotemporal cortex (not shown) where object recognition takes place. Model V1 interblobs contain both monocular and binocular cells. Binocular simple cells become disparity-sensitive by binocularly matching left and right scenic contours with the same contrast polarity in layer 3B before pooling opposite polarity responses at complex cells in layer 2/3A. Monocular and binocular boundary cells control filling-in of monocular 3D surfaces within V1 blobs. Closed boundaries can contain the filling-in process, and can send feedback to V1 interblobs that selectively strengthens the closed boundary components. Monocular and binocular V1 boundaries are pooled in V2. V2 pale stripes form long-range 3D perceptual groupings in which boundaries are completed and false matches inhibited to solve the correspondence problem. These completed boundaries form compartments in V2 thin stripes within which filling-in of monocular 3D surfaces occurs. Again, closed boundaries can contain the filling-in process and send surface-to-boundary feedback to enhance their boundaries, while also suppressing redundant boundaries at the same positions and farther depths. These boundaries and surfaces complete the representations of partially occluded objects. They do not generate visible percepts, but can be recognized by activating inferotemporal cortex. Visible surfaces in which figures are separated in depth from their backgrounds are formed in V4, where left and right eye surface signals are binocularly matched and pruned before filling-in a visible surface percept within enriched binocular boundaries from V2. V4 emits output signals that lead to recognition and grasping of unoccluded parts of opaque surfaces. See text for details.

As shown in Figure 5, both the boundary and surface streams receive illuminant-discounted signals from LGN cells with center-surround receptive fields, and both converge in V4, where visible 3D surfaces are consciously seen that are separated from their backgrounds. Model V2 and V4 also output to inferotemporal cortex (IT, not shown) where object recognition takes place. The 3D LAMINART model embodies several design constraints which are individually simple. However, when they work together to generate emergent perceptual properties, they can explain challenging and subtle percepts. Although the exposition below separates the BCP and SCP problems for clarity, boundary and surface stream interactions are needed to solve both problems. Model stages are mathematically described in the Appendix. Table 1 lists some of the anatomical and neurophysiological data that support the model.

	Connection in model (all in V1 unless otherwise noted)	Functional interpretation	Selected references
1	LGN→4	Strong, oriented LGN input	Blasdel and Lund (1983); Ferster *et al.* (1996, cat)
2	4→2/3 pyramidals	Feedforward of stimuli with bottom-up support	Fitzpatrick *et al.* (1985); Callaway and Wiser (1996)
3	2/3 pyramidals. →2/3 pyramidals	Long-range collinear integration along receptive field axes	Bosking *et al.* (1997, shrew); Schmidt *et al.* (1997, cat); Tucker and Katz (2003a, b, ferret)
4	2/3 pyramidals →2/3 inhibitory interneurons	Keep outward grouping subthreshold (bipole property)	McGuire *et al.* (1991); Hirsch and Gilbert (1991, cat); Tucker and Katz (2003a, b, ferret)
5	2/3 inhibitory interneurons →2/3 inhibitory interneurons	Normalize 2/3 inhibition (2-against-1 part of bipole property)	Tamas *et al.* (1998, cat); Tucker and Katz (2003a, b, ferret)
6	V1 2/3 pyramidals →V2 layer 4	Feedforward of V1 groupings into V2	Van Essen *et al.* (1986); Rockland and Virga (1990)
7	Presence of simple cells and binocular cells in layer 3B of V1	Contrast sensitivity in layer 3B and obligate property	Dow (1974); Hubel and Wiesel (1968); Poggio,(1972); Katz *et al.* (1989)
8	3B → 2/3 in V1 and the presence of Binocular and complex cells in layer 2/3	Pooling responses of layer 2/3 of both contrast polarity from layer 3B	Callway (1998); Poggio (1972)
9	Presence of cells in layer 3B and 2/3 that exclusively respond to binocular, not monocular stimulation	Obligate property	Poggio and Fischer (1977, rhesus); Smith et al. (1977); Poggio and Talbot (1981, rhesus); Poggio (1991)
10	Presence of monocular cells in layers 2 and 3	V1 monocular boundary formation	Poggio (1972); Hubel and Wiesel (1968)
11	V2 cells are mostly binocular	Model V2 cells (layer 4) input from both ocularities of monocular V1 cells (layer 2/3)	Hubel and Livingstone (1987); Roe and Ts'o (1997)
12	V2 cells are disparity sensitive	Depth detection in V2	Poggio and Fischer (1977, rhesus); von der Heydt et al. (2000); Peterhans, (1997)
13	No false matches in V2	Disparity filter in V2	Bakin et al., (2000)
14	Presence of false matches in V1	Depth propagation in model V1	Cumming and Parker (2000)
15	Presence of many complex cells in V2	Exclusive implementation of complex cells in the model V2	Hubel and Livingstone, (1987)
16	Interblob-blob connections in V1	V1 boundary-surface interaction and surface-to-boundary feedback	Rockland & Lund, 1983; Malach, 1992; Lund et al., 1993
17	Pale stripe-thin stripe connections in V2	V2 boundary-surface interaction and surface-to-boundary feedback	Rockland, 1985; Levitt et al., 1994; Lund et al., 1993
18	V2→IT	Recognizable copies of boundary and surface	Seltzer and Pandya, (1978); Baizer et al. (1991)

Table 1. Anatomical and physiological data that support the model.

The first model constraints concern how depth-selective boundaries are formed to solve the BCP.

V1 binocular filtering: contrast-specific fusion and contrast-invariant boundary. Several processing stages cooperate to ensure that each binocular match in the brain corresponds to the same object feature in the world. The first stage ensures that only edges in the left and right retinal images with the same contrast polarity can be binocularly matched (Howard and Rogers, 2002); that is, binocular fusion is *contrast-specific*, and obeys the *same-sign hypothesis*. Positionally disparate like-polarity monocular simple cells in layer 4 of V1 realize this property by conjointly activating like-polarity binocular simple cells in layer 3B of V1 (Figure 6).

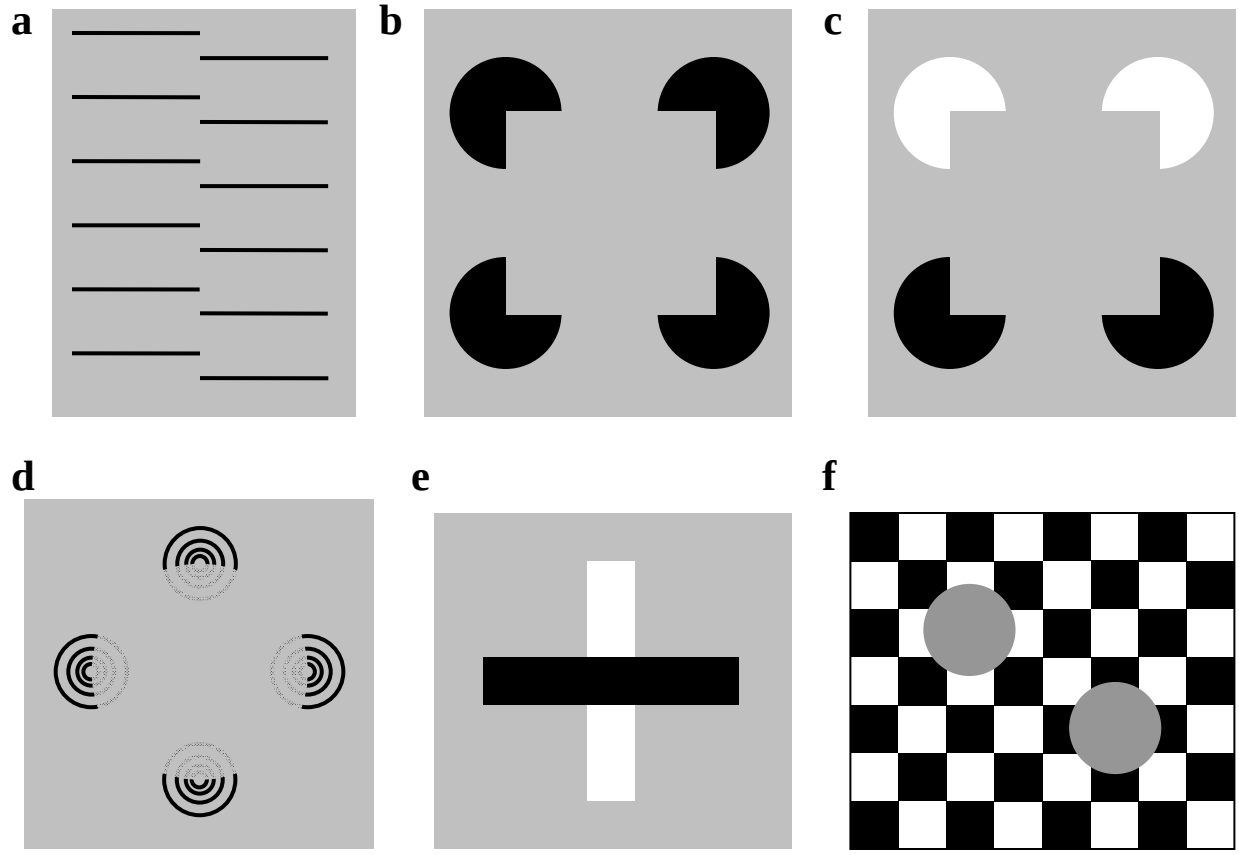


Figure 7. Examples of boundary completion and surface filling-in. (a) Offset grating: The offset black horizontal lines induce an amodal percept of a vertical boundary in the middle of the image that can be recognized even though it does not generate a visible brightness difference. (b) Kanizsa square: Four pacmen induce a percept of an illusory square that superimposes parts of four partially occluded disks. Visible brightness difference is also induced so that the square surface looks a bit brighter than its background. (c) Reverse-contrast Kanizsa square: When two pacmen reverse their contrast with regard to the background, an illusory square is still perceived, but no visible brightness difference is perceived between the square and its background. (d) Neon color spreading: Several circular lines with inner segments of weaker contrast than the outer segments induce a percept of a bright and transparent disk that

superimposes the less contrastive segments. (e) Partial occlusion: Two white vertical bars abutting the black horizontal bar induce a percept of a continuous white vertical bar that is amodally occluded behind an occluding black horizontal bar. (f) Polarity pooling: An object boundary is formed around each circular disk even though the contrast reverses polarity as the boundary is traversed.

Binocular boundaries are also *contrast-invariant*: Fused boundaries, such as the boundary around the gray disk in Figure 7f, can form around objects whose contrast polarity with respect to the background can reverse along their perimeters (Grossberg, 1994). Complex cells in layer 2/3 of V1 begin to realize this property by pooling outputs from binocular simple cells with opposite contrast-polarity selectivity at the same position and disparity (Figure 6). By pooling over both polarities, complex cells cannot tell if they are processing a light-dark or a dark-light polarity contrast, Hence, complex cells cannot represent the lightness or color of a boundary. In this sense, "all boundaries are invisible" (Grossberg, 1994). The brain hereby reconciles *contrast-specific* fusion with *contrast-invariant* object boundary formation by using at least three types of cortical V1 cells. The brain also preserves monocular information in V1 for reasons that are noted below.

V1 binocular filtering: Contrast magnitude constraint on binocular fusion. Preventing false binocular matches is also facilitated by restricting binocular fusion to only left and right eye signals that represent approximately the same magnitude of contrast (McKee, Bravo, Taylor, and Legge, 1994). Cortical cells that accomplish this property obey an *obligate property* (Poggio, 1991), which is realized when excitatory layer 4 inputs to layer 3B binocular simple cells also excite inhibitory interneurons, which inhibit each other as well as layer 3B binocular simple cells. Excitation and inhibition are hereby balanced to respond selectively to balanced contrast magnitudes (Figure 6) (Grossberg and Howe, 2003).

V2 encourages unique-matching using a disparity filter. Matching only the same contrast polarity at a similar contrast still allows many false binocular matches to occur. Some models try to remove them by imposing a *unique-matching rule*, which allows a feature from one retinal image to match at most one feature from the other (Grimson, 1981; Marr and Poggio, 1976). This rule fails in Panum's limiting case (Gillam, Blackburn, and Cook, 1995; Panum, 1858), where a bar presented to one eye can match two separate bars presented to the other eye. The brain needs more flexibility to resolve surface depth in situations, like DaVinci stereopsis, as discussed below, where monocular information is needed to see the correct percept.

The 3D LAMINART model encourages unique matches, but does not require them, using a *disparity filter* in V2 layer 2/3 (Figure 6) whereby active cells that share the same line-of-sight across different depths inhibit each other (Cao and Grossberg, 2005; Grossberg and Howe, 2003; McLoughlin and Grossberg, 1998). The disparity filter solves the correspondence problem, and explains Panum's limiting case (Cao and Grossberg, 2005; Grossberg and Howe, 2003). For example, in response to a dense RDS (Figure 1a), most false matches that survive processing in the model's V1 (Figure 1b) are suppressed in V2 by such line-of-sight inhibition (Figure 1d). This simulation result clarifies data showing that false matches occur in V1 but less readily in V2 (Bakin, Nakayama, and Gilbert, 2000; Cumming and Parker, 2000).

The disparity filter does not act alone. Two additional processes work with it (see Figures 5 and 6): (1) Perceptual grouping occurs in layer 2/3 of V2 pale stripes. It selects and completes the boundaries that control surface formation in V2 thin stripes and V4. (2) Surface-to-boundary feedback occurs from V2 thin stripes to pale stripes. It helps to ensure the consistency of

boundary and surface representations. How do complex cells in V1 layer 2/3 activate these V2 processes?

V2 combines monocular and binocular information to form depth percepts. An object's edge that is seen by one eye may be occluded and thus not seen in the other eye, as occurs during da Vinci stereopsis (Gillam, Blackburn, and Cook, 1995; Nakayama and Shimojo, 1990). Despite a lack of binocular information, the monocularly viewed region has a definite depth conferred to it by binocularly viewed parts of the scene. The brain can thus utilize monocular information to build up seamless 3D percepts of the world (Grossberg, 1994). The model assumes that monocular and binocular pathways are separated in the V1 interblobs, but that layer 4 of V2 pale stripes combines the monocular and binocular outputs from V1 (Figure 6). In particular, monocular boundary outputs from V1 are added to V2 binocular cells along their lines-of-sight at all depths. The disparity filter, with the help of surface-to-boundary feedback, automatically eliminates most of monocular boundary copies at the incorrect depths, thus conferring correct depths to initially depth-ambiguous monocular boundaries (Cao and Grossberg, 2005; Grossberg and Howe, 2003). This prediction has been tested with positive results (Yazdanbakhsh and Watanabe, 2004).

V2 perceptual grouping completes boundaries and eliminates false matches. When contours of an object are incomplete (e.g., Figures 7a–e), boundary completion can link them. This process facilitates recognition of partially occluded objects (Figure 7e). The model predicts how boundary completion is carried out by pyramidal cells in layer 2/3 of V2 pale stripes (Figure 6) (Grossberg and Raizada, 2000), consistent with neurophysiological experiments (von der Heydt, Peterhans, and Baumgartner, 1984). Model cells whose connections are approximately collinear along their common preferred orientation excite each other via long-range horizontal connections. These connections also activate inhibitory interneurons that inhibit each other and nearby pyramidal cells via short-range disynaptic inhibition (Figure 6). This balance of excitation and inhibition at target cells implements the *bipole property* (Cohen and Grossberg, 1984; Grossberg and Mingolla, 1985), whereby a cell is activated when its receptive field simultaneously receives horizontal inputs on two sides from approximately collinear positions and aligned orientation preferences, but not by an input from only one side. Perceptual groupings can hereby form inwardly, as in Figures 7a – e, but not outwardly.

A subset of these inhibitory interneurons is predicted to realize the disparity filter via line-of-sight inhibition across different depths. This hypothesis puts the disparity filter within layer 2/3 of V2 pale stripes as part of the perceptual grouping process. The elimination of "false matches" and "weak and incorrect groupings" are hereby both predicted to be achieved by the 3D perceptual grouping process. This prediction explains how an emergent perceptual grouping can sometimes override local disparities in determining perceived depth (Ramachandran and Nelson, 1976).

V1 and V2 amodal surface representations. The above constraints all concern how the brain constructs a 3D boundary representation of an object. However, "all boundaries are invisible". Visibility is a property of surfaces (Grossberg, 1994). Unlike in the boundary stream, visible contrast and color information is preserved and enhanced by opponent and double-opponent cells in the surface stream. Indeed, most cells in V1 blobs and V2 thin stripes are insensitive to orientation, but sensitive to color (Hubel and Livingstone, 1987; Ts'o, Roe, and Gilberg, 2001). Discounting the illuminant begins at an early surface-processing stage (e.g., retina) to suppress lightness and color signals that vary slowly across space (Grossberg and Hong, 2006; Grossberg and Todorović, 1988; Land and McCann, 1971). A surface *filling-in process*

spreads lightness and color signals that survive the illuminant discounting process across space until they encounter a boundary or attenuate due to their spread (Grossberg, 1987; Grossberg, 1994; Grossberg and Hong, 2006; Grossberg and Mingolla, 1985; Grossberg and Todorović, 1988). Surface filling-in is predicted to occur in V1, V2, and V4 with different functional consequences that the model explains below. The model predicts that visible visual percepts are associated with filled-in V4 surfaces, whereas filled-in surfaces within V1 blobs and V2 thin stripes are invisible (amodal, or without visible features) under normal circumstances. The current work predicts a new functional role for V1 surface filling-in and surface-to-boundary feedback, as noted below.

Only closed, or almost closed, 3D boundaries can contain the filling-in of lightness and color signals within the enclosed region, instead of allowing them dissipate through boundary gaps (Grossberg, 1994). This fact enables the surface system to help the boundary system resolve a problem that it cannot solve on its own. Indeed, depth-ambiguous horizontal boundaries are replicated along their line-of-sight at multiple depths in layer 4 of V2. Only at depths when horizontal boundaries complete a closed boundary with depth-selective (say, vertical) binocular boundaries can filling-in be contained. These closed boundaries can carry out selective 3D *surface capture* through a boundary-surface feedback interaction, as explained below. The prediction that vertical and horizontal boundaries are processed differently to determine depth has been successfully tested (Yazdanbakhsh and Watanabe, 2004).

Filling-in of a closed boundary within a V1 blob implicitly tags the monocular horizontal boundaries that form the closed boundary with a particular depth. Surface-to-boundary blob-to-interblob feedback signals detect the contrast around the filled-in surface and send excitatory signals to the corresponding V1 boundaries, thereby strengthening them. These strengthened horizontal boundaries are favored during disparity filter boundary competition in V2 pale stripes, thereby selecting the favored horizontal boundaries and suppressing the redundant ones.

These selected boundaries then help to determine what regions successfully fill-in within V2 thin stripes, because successful filling-in within a V2 thin stripe is also restricted to closed boundary regions. Surface-to-boundary feedback from thin stripes to pale stripes strengthens the corresponding boundaries, and inhibits redundant boundaries at the same positions but farther depths. This allows partially occluded boundaries (Figure 8a) to be completed in depth behind occluders (Figure 8b), and to thereby fill-in their closed boundary in the thin stripes (Figure 8c). V2 thin stripes are hereby predicted to form complete surface representations for both occluding and occluded objects). These complete boundary and surface representations input to object recognition processes in inferotemporal cortex.

Although partially occluded, opaque objects, as in Figures 7e-f and 8a, can be recognized, their occluded parts cannot be seen. Otherwise every opaque occluder would be perceived as transparent. The model predicts that surface representations in V2 thin stripes are *amodal surfaces*, which are suitable for recognition, but are not seen as visible percepts.

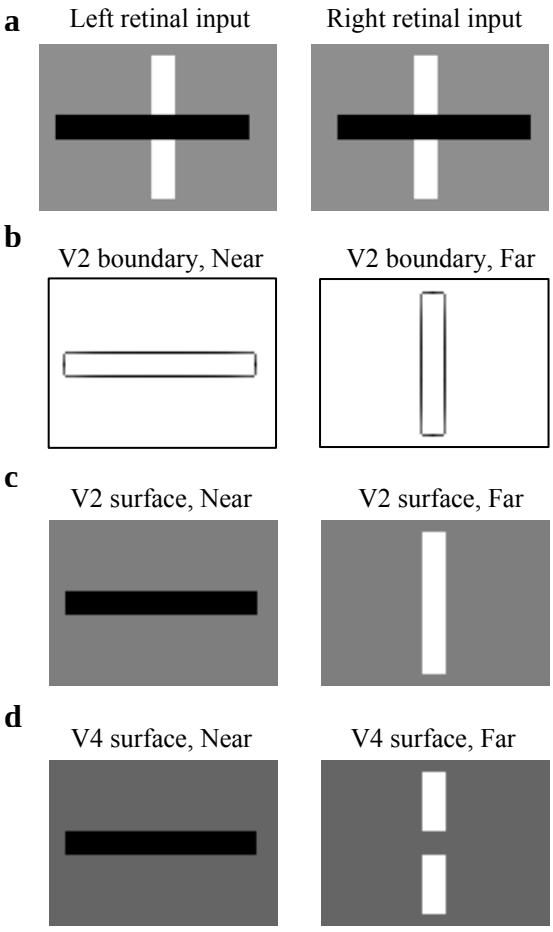


Figure 8. 3D figure-ground separation and amodal completion: (a) Retinal inputs. (b) Boundaries at bipole cells in layer 2/3 of model V2 pale stripes. (c) Filled-in amodal surface representation in model V2 thin stripes. In particular, a complete surface of the occluded object is captured and filled-in at the far depth. It enables inferotemporal cortex to recognize the occluded surface color, even though that color is amodal, or perceptually invisible. (d) Visible, or modal, surface representation in V4.. See text for details.

V4 visible surface representations: Boundary enrichment and surface pruning. The surface representations in V4 are visible, or modal, surfaces. Why are only the unoccluded surfaces of opaque objects (e.g., Figure 8a) seen in such percepts? The model predicts that V2 boundaries at a certain depth project to V4 boundaries at their own depth and at all farther depths (*boundary enrichment*). Surface signals project to V4 along two pathways (Figure 5): Monocular illuminant-discounted surface signals directly excite V4. Successfully filled-in V2 surfaces at a certain depth inhibit surface features within V4 at all farther depths at their own positions (*surface pruning*) (Grossberg, 1994). In V4, the pruned surface features activate filling-in of visible 3D surfaces within the compartments defined by the enriched 3D boundaries. As a result,

the positions at depths behind an opaque occluder cannot fill-in, hence look opaque (Figure 8d). These same mechanisms also explain when transparent surface percepts can arise due to filling-in at the same positions and more than one depth (Grossberg and Yazdanbakhsh, 2005).

Results

As noted in the Introduction, the current refinement of the 3D LAMINART model explains three types of stereogram surface percepts: sparse and dense stereograms, and stereograms which generate emergent percepts of occluded objects that can be completed behind their occluders.

Explaining dense RDS percepts. When the images in Figure 1a are fused, two textured "L" shaped bars are perceived to be floating above a textured background. Furthermore, the upright "L" shaped bar on the left side is perceived to be farther than the reversed one on the right side. Figure 1b shows the boundaries that are represented by the activation patterns of model V1 binocular complex cells. There are many false matches on every depth plane. Figure 1c illustrates the role of surface-to-boundary feedback from the V1 blobs to interblobs: There are many more spurious horizontal boundaries than in Figure 1b when blob-to-interblob feedback is prevented. Figure 1d shows the boundaries that are formed in in layer 2/3 of V2 pale stripes after the disparity filter acts to suppress spurious boundaries. Most false binocular matches are suppressed, while correct ones are preserved, thereby solving the correspondence problem. Figure 1e shows the filled-in visible surfaces that are represented by the activation patterns of model V4 binocular surface cells (see Figure 5). Filling-in is controlled by the boundaries in Figure 1d. Because the "L" boundaries are disjoint, boundary pruning in V4 plays no role in the final percept.

Explaining sparse RDS percepts. The sparse stereogram in Figure 2a is designed so that, when the two images are successfully fused, a white square dotted with black is perceived to be floating over a white background that is also dotted with black. The square is perceived to be slightly whiter than the background. The critical question is how the large feature-absent spaces between the dots are perceived in the correct depths. Figure 2b shows the boundaries represented by the activation patterns of model cells in layer 2/3 of model V2 pale stripes. Illusory boundaries formed by a bipole grouping network in layer 2/3 of the V2 pale stripes connect the spatially sparse dots and enable the outline of a big square to form in the boundary stream. These emergent boundaries play a key role in generating the percept. Figure 2c shows that a square-shaped surface with black dots is captured in the V2 thin stripes through surface filling-in at the near depth, and separated from the background surface at the far depth. The illusory boundaries contain the filling-in of whiteness within the square, and make the white square recognizable in the surface stream. White can fill in except within the inducing dots, which create barriers to filling-in. The dots fill-in black within their own boundaries. Figure 2d shows the visible 3D surfaces that are represented by the activation patterns of V4 cells: At the fixation depth plane, the illusory boundaries contain the filling-in of white outside the black dots and within the square, while the dots are filled-in with blackness. Thus both the dots and the whole square surface are visible at the fixation depth. At the far depth, the region occluded by the central square formed at the fixation depth is prohibited from being filled-in (the gray level there corresponds to zero neural activity) by boundary enrichment and surface pruning, which explain why the square looks opaque instead of transparent. The rest of the background surface is filled-in with white that makes it visible at the far depth. Once again, the dots fill in black within their own boundaries.

Explaining amodal completion of partially occluded RDS. In Figure 4a, a vertical textured bar at the near depth and two textured squares at the fixation depth are defined by binocular disparity (Figures 3b and 3c). In addition, the two textured squares induce a percept of amodal completion of a horizontal bar behind the vertical textured bar (Figure 3d). Explaining this percept requires *multiple-scale* boundary processing (Andrews and Pollen, 1979; Foster, Gaska, Nagler, and Pollen, 1985; Tootell, Silverman, and De Valois, 1981). Two boundary scales are simulated here, for simplicity.

Small-scale boundaries within the V2 pale stripes (Figure 3b) control the amodal and visible surface percepts that fill in the texture compartments within the V2 thin stripes and V4, respectively. Within V2, small-scale near boundaries of the vertical bar inhibit the vertical bar's horizontal and vertical boundaries at the fixation plane (middle column, Figure 3b). This is achieved in V2 by near-to-far inhibition surface-to-boundary feedback and line-of-sight inhibition in the disparity filter. Figure 3e shows the visible surface percept that is generated by the boundaries in Figure 3b.

How does amodal completion of the partially occluded textured bar occur in the fixation plane? How does it overcome the small-scale boundary inhibition of the fixation planej? How does it span a gap that is so much larger than the small-scale boundary size? We propose that this is achieved by large-scale boundaries.

Large-scale boundaries register only the outlines of a small-scale textured figure. Their responses within the textured surface are suppressed because the excitatory and inhibitory halves of large-scale simple cell receptive fields receive approximately equal inputs, which cancel each other out (Figure 4c). Large-scale occluder boundaries at the near depth inhibit the large-scale boundaries of the occluded object at its own positions in the fixation plane. This is achieved by near-to-far inhibition in the V2 surface-to-boundary feedback. As a result, the right vertical boundary of the left square, and the left vertical boundary of the right square are inhibited. Boundary completion of the horizontal contours of the occluded object can then occur in layer 2/3 of model V2 pale stripes (Figure 3d, middle panel), without interference from the vertical boundaries of the occluded object. The large-scale boundaries can also be completed over a longer gap, thereby creating a completed amodal percept of a partially occluded horizontal bar.

Discussion

From complementary boundary and surface rules to consistent percepts. The 3D LAMINART model predicts functions of reciprocal connections between V1 blobs and V1 interblobs (Malach, 1992; Rockland and Lund, 1983), and between V2 thin stripes and V2 pale stripes (Levitt, Yoshioka, and Lund, 1994). These connections are predicted to realize interactions between the boundary and surface streams: Boundaries control surface filling-in during the 3D surface capture process. Successfully filled-in regions within the surface stream send *surface-to-boundary feedback* to the boundary system. The surface system does this by sensing whether or not a surface region is filled-in by using contrast-sensitive output circuits that can detect where the bounding contours of the filled-in region occur. Such circuits can distinguish whether a surface region contains its internal lightness or color within a closed boundary, or allows it to dissipate if it does not possess a closed boundary. These contrast-sensitive output signals are realized by cells that interact via a shunting on-center off-surround network (Grossberg, 1980) that operates within disparity and across position within the surface system.

The outputs from the surface stream to the boundary stream strengthen, and thereby confirm, the boundaries that surround the successfully filled-in surface regions. At the same time

they inhibit, or prune, redundant boundaries at the same positions and farther depths (Grossberg, 1994). This strengthening-and-pruning operation is accomplished by contrast-sensitive shunting on-center off-surround networks (Grossberg, 1980) that operate across disparity and within position in the boundary system. Such networks respond only at the bounding contours of successfully filled-in surface regions that are surrounded by closed, or almost-closed, boundaries.

A primary function of the feedback is to ensure the *consistency* between boundaries and surfaces (Grossberg, 1994; Grossberg, 2000). This is needed because boundaries and surfaces obey *complementary* computational rules (Figure 5). The same feedback mechanism also helps to determine the correct depth assignments for horizontal and monocular boundaries, which initially are depth-ambiguous. Surface-to-boundary feedback from V1 blobs to interblors is a new feature of the current model, one that helps to eliminate many otherwise depth-ambiguous horizontal boundaries. Surface-to-boundary feedback from V2 thin stripes to pale stripes confirms border ownership of the occluding figures, and thereby contributes to include 3D figure-ground separation. 3D LAMINART hereby clarifies how brain evolution could evolve figure-ground separation. The model predicts that the blob and interblob streams exchange a simple type of feedback signals that select a consistent percept despite their complementary computational rules. Figure-ground separation is an emergent property of this selection process. A neurophysiological who records from V1 or V2 boundary cells will thus often record a mixture of boundary and surface properties, including polarity-sensitive surface properties, despite the fact that boundaries pool signals over opposite contrast polarities.

Every stage in the model, including its laminar interpretation, is supported by neurophysiological and anatomical data. Such data are reviewed elsewhere (Grossberg, 1994; Grossberg and Howe, 2003; Grossberg and Raizada, 2000). The model does not include cortical areas V3, V3A, and MT, that are known to be involved in depth perception. These areas are not required to simulate the present data. They may be required when the model is extended to include motion perception and eye movement control. See Fazl, Grossberg, and Mingolla (2007) for an interpretation of V3A in eye movement control.

The model is minimally complex in the sense that each of its six interacting networks, V1 monocular boundaries, V1 binocular boundaries, V2 binocular boundaries, V1 monocular surfaces, V2 monocular surfaces, and V4 binocular surfaces (Figure 6) is essential to explain key aspects of targeted data: The V1 binocular boundaries network is needed to explain the contrast-polarity constraint and the contrast magnitude constraint observed in stereoscopic fusion (Smallman and McKee, 1995). The V1 monocular boundaries network is needed to explain many data involved with monocular-binocular combination, including da Vinci stereopsis (Nakayama and Shimojo, 1990), dichoptic masking (Smallman and McKee, 1995), contrast variations of the correspondence problem (Smallman and McKee, 1995), and stereopsis with opposite contrast stimuli (Howe and Watanabe, 2003). Indeed, the current refinement of 3D LAMINART is capable of simulating all the data previously simulated by the LAMINART family of models.

The V2 boundaries network is needed to solve the *correspondence problem*, the *monocular-binocular interface problem*, and the *figure-ground problem*. The correspondence problem arises because V1 sometimes incorrectly fuses contours that belong to different objects. The monocular-binocular interface problem is caused because the V1 monocular boundaries, not having a definite depth association, are initially added to all depth planes. The figure-ground problem arises because, when partial occlusion occurs, shared borders are exclusively owned by the occluding object and visible parts of the occluded object need to be amodally grouped. The

V1 monocular surfaces network is needed to offset the fixation plane bias (Cao and Grossberg, 2005; Grossberg and Howe, 2003), which would otherwise always favor horizontal boundaries in the fixation plane when line-of-sight competition takes place in the disparity filter. Surface-to-boundary feedback overcomes this bias by enhancing all closed boundaries that contain the filling-in of surface-feature signals. The V2 monocular surfaces network is needed to explain the recognizable but amodal surfaces in figure-ground perception, and the surface-to-boundary feedback that it generates is essential in the determination of border ownership. The binocular surface network in V4 is needed to explain visible surface percepts. Finally, the multiple-scale processing for boundaries is necessary to explain figure-ground percepts for textured figures that include emergent occlusion.

APPENDIX

Mathematical Framework and Simulation Methods

The simulations of the model were done with the Matlab software package. All simulations used the same model parameters. Cells in Retina/LGN and each cortical layer are described by first-order differential equations. Except for cells in V2 layer 2/3 and filling-in processes within V1 blobs, V2 thin stripes and V4, the equilibrium states are calculated. For cells in V2 layer 2/3 and filling-in processes, Euler's method is used, and the results are verified by multiple step sizes. After surface filling-in under reaches equilibrium, surface-to-boundary feedback is calculated at the boundary stream to render a new equilibrium of the boundaries. Generally after three loops, the boundary and surface streams become approximately consistent with each other and reached relative equilibrium. Up to 35,328,000 cells were recruited in one simulation.

Membrane equations. Model LGN cells and model boundary cells in V1 and V2 were simulated by membrane, or shunting, equations consisting of interacting neurons, each with a single voltage compartment whose membrane potential, *V(t)*, that obeys:

$$c_m \frac{dV(t)}{dt} = -[V(t) - e_{leak}]\gamma_{leak} - [V(t) - e_{excit}]\gamma_{excit}(t) - [V(t) - e_{inhib}]\gamma_{inhib}(t) \qquad (1)$$

In equation (1), c_m is the membrane capacitance. The γ_{leak} term is a constant leakage conductance. The e terms represent reversal potentials of the leakage, excitatory, and the inhibitory channels, respectively. Time-varying conductances $\gamma_{excit}(t)$ and $\gamma_{inhib}(t)$ represent, respectively, the total inputs from the excitatory and inhibitory neurons that make synapses onto the cell, as determined by the model architecture shown in Figures 5 and 6. At equilibrium, the above equation can be written as $V = \frac{e_{leak}\gamma_{leak} + e_{excit}\gamma_{excit} + e_{inhib}\gamma_{inhib}}{\gamma_{leak} + \gamma_{excit} + \gamma_{inhib}}$. Thus, increases in the excitatory and inhibitory conductances depolarize and hyperpolarize the membrane potential, respectively, as shown by the numerator of this term, and all the conductances contribute to divisive normalization of the membrane potential, as shown by the denominator. This divisive effect includes the special case of pure "shunting" inhibition when the reversal potential of the inhibitory channel is close to the cell's resting potential (Borg-Graham, Monier, and Fregnac, 1998; Grossberg, 1973). In the following examples of shunting inhibition, the reversal potentials are: $e_{excit} > 0$, $e_{inhib} = e_{leak} = 0$, and the membrane conductance $c_m = 1$, except where otherwise indicated. Then equation (1) reduces to:

$$\frac{dV(t)}{dt} = -\varepsilon V(t) + [\alpha - V(t)]\gamma_{excit}(t) - V(t)\gamma_{inhib}(t), \qquad (2)$$

where ε is a constant (10^{-5}) that represents γ_{leak}, and α is a constant (10) that represents e_{excit}. ε and α are constants with the above values all through this Appendix. When the excitatory and inhibitory inputs γ_{excit} and γ_{inhib} are time-invariant, the membrane potential *V* at equilibrium is:

$$V = \frac{\alpha\gamma_{excit}}{\varepsilon + \gamma_{excit} + \gamma_{inhib}} \tag{3}$$

Notation. Outputs of model cells are represented by capital letters. They may be supplemented by upper and lower indices. Some common indices are summarized here. For example, consider the output, $S_{(i\pm e)jkd}^{B/L/R,s/l,+/-}$, of a simple cell in layer 3B of V1 interblob: For the upper indices, *B/L/R* represents a Binocular, or Left eye, or Right eye input source; *s/l* represents a small or large scale receptive field; and +/- represents a light-to-dark or dark-to-light contrast-polarity. For the lower indices, *i,j,* are position indices, with *i* representing the horizontal coordinate (dummy index *p* is also used), and *j* representing the vertical coordinate (dummy index *q* is also used); *k* represents the model cell orientation tuning (two orientations, horizontal and vertical, are simulated); *d* represents disparity or depth (dummy index *d'* is also used), where a smaller index represents a nearer depth and *d=1* represents the nearest depth (in this article, three depths are simulated); *e* represents the allelotropic shift that depends on the disparity *d* (dummy index *e'* is also used, which depends on disparity *d'*). The relationship between allelotropic shift and disparity is defined for the near disparity by a shift e = -16, for the fixation disparity by a shift e = 0, and for the far disparity by a shift e = +16. In another example, output of a cell in either V1 blobs, V2 thin stripes, or V4 is represented by $Z_{ijd}^{B/L/R,on/off,1/2/4}$: Here, *L/R* represents that the cell belongs to the Left or Right monocular Filling-In Domain (FIDO) in V1 blobs or V2 thin stripes, while *B* represents that the cell belongs to the V4 binocular FIDO; *on/off* represents that the cell belongs to the ON or OFF FIDO (for Retina/LGN cells, *on/off* represents that the cell has an on-center off-surround or off-center on-surround receptive field, respectively); and *1/2/4* represents that the cell belongs to either V1 blobs, V2 thin stripes or V4. In some equations operator "$[\ ]^+$" is used to represent a half-wave rectifying output signal threshold; e.g., $[x]^+ = \max(x,0)$.

Filling-in and contrast-sensitive outputs. Filling-in processes occur in the FIDOs of V1 blobs, V2 thin stripes and V4. A general process of filling-in and generation of surface-to-boundary feedback is described herein. This process clarifies how boundary signals Y_{ijkd} control filling-in of surface-feature inputs $X_{ijd}^{on/off}$ to generate filled-in surface activities $Z_{ijd}^{on/off}$ and contrast-sensitive outputs $F_{ijd}^{on/off}$. The shunting lateral inhibition in LGN ensures that surface-feature inputs occur only at contrast changes, and that these inputs estimate the ratio of the luminances on either side of the border. The filling-in process is monocular when the left and right monocular FIDOs fill-in separately; e.g., in V1 blobs and V2 thin stripes; and binocular in V4 where the binocular FIDO receives inputs from both eyes via LGN. In addition, V1 blobs and V2 thin stripes send contrast-sensitive outputs to V1 interblobs and V2 pale stripes, respectively, to initiate surface-to-boundary feedback. Contrast-sensitive outputs from V2 thin stripes are also sent to V4, to initiate surface pruning.

Following Grossberg and Todorović (1988), the filling-in process is modeled by a nearest-neighbor diffusion equation in both ON and OFF cells:

$$\varphi_z \frac{d}{dt} z_{ijd}^{on/off} = -z_{ijd}^{on/off} + \sum_{(p,q)\in n(i,j)} (z_{pqd}^{on/off} - z_{ijd}^{on/off}) r_{pqijd} + X_{ij}^{on/off}, \tag{4}$$

where positive rate parameter $\varphi_z << 1$ ensures that filling-in is fast; $z_{ijd}^{on/off}$ represent the activities of the model cells; $X_{ij}^{on/off}$ represents the surface-feature inputs from LGN; and the nearest neighbors $n(i, j)$ of position *(i,j)* are defined by:

$$n(i, j) = \{(i-1, j), (i+1, j), (i, j-1), (i, j+1)\} \tag{5}$$

The gating coefficients r_{pqijd} in equation (4) control the diffusion rate between positions *(i,j)* and *(p,q)* at the depth plane corresponding to disparity d, by the boundary signals Y_{ijkd}. In particular, when $(p, q) \in n(i, j)$:

$$r_{pqijd} = \frac{\mu}{1 + \nu \sum_{(l,m)\in\omega(p,q,i,j)} \sum_{k} Y_{lmkd}}, \tag{6}$$

where constants μ and ν (1000, 10000) represent the diffusion rate and the strength of the gating inhibition, while the neighboring area $\omega(p, q, i, j)$ from which a boundary signal can block the filling-in of the surface-feature signals is defined as:

$$\omega(\mathrm{p,q,i,j}) = \begin{cases} \{(i-0.5, j-0.5), (i-0.5, j+0.5)\}, \text{when (p, q)} = \text{(i - 1, j)}, \\ \{(i+0.5, j-0.5), (i+0.5, j+0.5)\}, \text{when (p, q)} = \text{(i + 1, j)}, \\ \{(i-0.5, j-0.5), (i+0.5, j-0.5)\}, \text{when (p, q)} = \text{(i, j - 1)}, \\ \{(i-0.5, j+0.5), (i+0.5, j+0.5)\}, \text{when (p, q)} = \text{(i, j + 1)}. \end{cases} \tag{7}$$

The boundary lattice is offset by [0.5, 0.5] relative to the feature/surface lattice, corresponding to the idea that these two processing streams are spatially displaced with respect to one another in the cortical map. In the *Matlab* implementation, however, such half-pixel indexes in the boundary system are readjusted to integral indexes by subtracting (0.5, 0.5), so that the index of the most up-left boundary element is (1, 1).

To ease the computational burden, equation (4) is solved at steady-state iteratively until equilibrium is achieved, using the steady-state equation derived from equation (4):

$$z_{ijd}^{on/off} = \frac{X_{ij}^{on/off} + \sum_{p,q\in n(i,j)} z_{ijd}^{on/off} r_{pqijd}}{1 + \sum_{p,q\in n(i,j)} r_{pqijd}} \tag{8}$$

The surface outputs of the FIDO, $Z_{ijd}^{on/off}$, are mediated through an opponent structure, in which cells on the same position in the ON-channel and the OFF-channel inhibit each other. Since $X_{ij}^{on/off}$ already involve surround inhibition, this defines a double-opponent network. The double-opponent operation obeys:

$$Z_{ijd}^{on} = \left[z_{ijd}^{on} - z_{ijd}^{off}\right]^{+} \tag{9}$$

and

$$Z_{ijd}^{off} = \left[z_{ijd}^{off} - z_{ijd}^{on}\right]^{+} \tag{10}$$

The contrast-sensitive cells in the FIDO are sensitive to the contrasts between the successfully filled-in regions. Their outputs, $F_{ijkd}^{s/l}$, obey:

$$F_{ijkd}^{s/l} = \left[\left(\left[\sum_{p,q} b_{pqk}^{s/l} Z_{i+p,j+q,d}^{on} \right]^+ + \left[\sum_{p,q} -b_{pqk}^{s/l} Z_{i+p,j+q,d}^{off} \right]^+ + \left[\sum_{p,q} -b_{pqk}^{s/l} Z_{i+p,j+q,d}^{on} \right]^+ + \left[\sum_{p,q} b_{pqk}^{s/l} Z_{i+p,j+q,d}^{off} \right]^+ \right) - \theta_f^{s/l} \right]^+, \tag{11}$$

where thresholds (θ_f^s , θ_f^l) are (0.1, 7), and $b_{pqk}^{s/l}$ are oriented Gabor kernels that obey:

$$b_{pqk}^{s/l} = \phi_b \sin\left(\frac{2\pi \cdot \xi_k}{\tau} \right) \exp\left[-\frac{1}{2} \left(\frac{p^2}{\sigma_p^2} + \frac{q^2}{\sigma_q^2} \right) \right], \tag{12}$$

whose kernel parameters (ϕ_b , τ , σ_p , σ_q) are constants that represent the amplitude and dimensions of the kernel, which are set to (4.4, 3, 0.6, 0.6) for the small scale, and (0.01, 120, 48, 48) for the large scale, respectively; $\xi_k = q$ for cells that are tuned to the vertical orientation; and $\xi_k = p$ for cells that are tuned to the horizontal orientation.

Equations of Model Processing Stages

The equations of model cells in the simplified retina/LGN and for each layer of V1, V2 and V4 cells are given below:

Retina/LGN. The Retina/LGN cells obey membrane, or shunting, equations. There are two types of retina/LGN cells: ON cells have circularly symmetric on-center, off-surround receptive fields, while OFF cells have circularly symmetric off-center, on-surround receptive fields. When these fields are approximately balanced, the network discounts the illuminant and contrast-normalizes its cell responses (Grossberg & Todorović, 1988). Left and right monocular pathways are separated from each other in retina/LGN. At each location, ON-cells and OFF-cells inhibit each other and form a double-opponent pair. The retina/LGN ON-cell membrane potentials, $x_{ij}^{L/R,on}$, obey the shunting equation:

$$\frac{dx_{ij}^{L/R,on}}{dt} = -\varepsilon x_{ij}^{L/R,on} + (\alpha - x_{ij}^{L/R,on}) I_{ij}^{L/R} - x_{ij}^{L/R,on} \sum_{p \neq i, q \neq j} g_{pqij} I_{pq}^{L/R}, \tag{13}$$

where $I_{ij}^{L/R}$ are the luminances from the left (L) or right (R) retinal image; and g_{pqij} is a Gaussian kernel that represents the inhibitory off-surround:

$$g_{pqij} = \phi_g \exp\left[-\frac{(p-i)^2 + (q-j)^2}{2\sigma_g} \right], \tag{14}$$

where ϕ_g is a constant (0.076) that scales the connection from nearby retinal cells, and σ_g is a constant (1.5) that controls kernel size.

The Retina/LGN OFF-cell membrane potentials, $x_{ij}^{L/R,off}$, obey the equation:

$$\frac{dx_{ij}^{L/R,off}}{dt} = -\varepsilon x_{ij}^{L/R,off} + (\alpha - x_{ij}^{L/R,off}) \sum_{p \neq i, q \neq j} g_{pqij} I_{pq}^{L/R} - x_{ij}^{L/R,off} I_{ij}^{L/R} \tag{15}$$

The double-opponent outputs obey the equations:

$$X_{ij}^{L/R,on} = \left[x_{ij}^{L/R,on} - x_{ij}^{L/R,off} \right]^+ \tag{16}$$

and

$$X_{ij}^{L/R,off} = \left[x_{ij}^{L/R,off} - x_{ij}^{L/R,on} \right]^+ \tag{17}$$

V1 Layer 4 simple cells. Model V1 layer 4 monocular simple cells are sensitive to either dark-light or light-dark contrast polarity, but not both, depending on their receptive field structure. Both small-scale and large-scale simple cell receptive fields are modeled. At steady-state, the outputs, $S_{ijk}^{L/R,s/l,+/-}$ (see Section A2 for the index notation), of simple cells are given by:

$$S_{ijk}^{L/R,s/l,+} = \left[\left[\sum_{p,q} b_{pqk}^{s/l} X_{i+p,j+q}^{L/R,on} \right]^+ + \left[\sum_{p,q} -b_{pqk}^{s/l} X_{i+p,j+q}^{L/R,off} \right]^+ - \theta_s^{s/l} \right]^+ \tag{18}$$

and

$$S_{ijk}^{L/R,s/l,-} = \left[\left[\sum_{p,q} -b_{pqk}^{s/l} X_{i+p,j+q}^{L/R,on} \right]^+ + \left[\sum_{p,q} b_{pqk}^{s/l} X_{i+p,j+q}^{L/R,off} \right]^+ - \theta_s^{s/l} \right]^+, \tag{19}$$

where $X_{ij}^{L/R,on/off}$ are defined by equations (16) and (17); (θ_s^s, θ_s^l) are the output threshold (7, 1); and $b_{pqk}^{s/l}$ are oriented Gabor kernels that are defined in equation (12), with kernel parameters (ϕ_b, τ, σ_p, σ_q) set to (4.4, 3, 0.6, 0.6) for the small scale, and (0.1, 30, 12, 12) for the large scale, respectively.

V1 Layer 3B monocular simple cells. The membrane potentials, $B_{ijk}^{L/R,s/l,+/-}$, of the layer 3B monocular cells obey at equilibrium:

$$B_{ijk}^{L/R,s/l,+/-} = 2 \cdot S_{ijk}^{L/R,s/l,+/-}, \tag{20}$$

where $S_{ijk}^{L/R,s/l,+/-}$ are the simple cells in V1 layer 4 defined by equations (18) and (19), and the multiplicative factor of 2 compensates for the fact that the monocular simple cells receive inputs from only one eye, whereas the binocular simple cells, discussed in the next section, receive input from both eyes.

V1 layer 3B binocular simple cells. Layer 3B binocular simple cells have membrane potentials $B_{ijkd}^{B,s/l,+/-}$. They receive excitatory inputs *S* from layer 4 simple cells and inhibitory inputs *Q* from the layer 3B inhibitory interneurons that correspond to the same position, orientation, and disparity. At equilibrium:

$$B_{ijkd}^{B,s/l,+/-} = \frac{1}{\gamma_1} \Big[S_{(i+e)jk}^{L,s/l,+/-} + S_{(i-e)jk}^{R,s/l,+/-} - \rho_1 \left([Q_{ijkd}^{L,s/l,+/-}]^+ + [Q_{ijkd}^{L,s/l,-/+}]^+ + [Q_{ijkd}^{R,s/l,+/-}]^+ + [Q_{ijkd}^{R,s/l,-/+}]^+ \right) \Big], \tag{21}$$

where γ_1 and ρ_1 are constants (0.29, 5) that represent the rate of decay of the membrane potential and the strength of the inhibition, respectively. Under mild constraints on these parameters, the binocular cells act like the "obligate cells" of Poggio (1991), responding only when their left and right inputs are approximately equal in magnitude. Equation (21) was solved at equilibrium, as justified by the theorem proved in Grossberg and Howe (2003), to speed up the simulations.

V1 layer 3B inhibitory cells. Layer 3B inhibitory cells receive excitatory inputs from layer 4 simple cells and inhibitory inputs from all other inhibitory cells that correspond to the same position, orientation, and disparity. Their cell membrane potentials, $Q_{ijkd}^{L/R,s/l,+/-}$, are determined at equilibrium by the equations:

$$Q_{ijkd}^{L,s/l,+/-} = \frac{1}{\gamma_2}\left[S_{(i+e)jk}^{L,s/l,+/-} - \rho_2\left([Q_{ijkd}^{R,s/l,+/-}]^+ + [Q_{ijkd}^{L,s/l,-/+}]^+ + [Q_{ijkd}^{R,s/l,-/+}]^+\right)\right] \quad (22)$$

and

$$Q_{ijkd}^{R,s/l,+/-} = \frac{1}{\gamma_2}\left[S_{(i-e)jk}^{R,s/l,+/-} - \rho_2\left([Q_{ijkd}^{L,s/l,+/-}]^+ + [Q_{ijkd}^{L,s/l,-/+}]^+ + [Q_{ijkd}^{R,s/l,-/+}]^+\right)\right], \quad (23)$$

where γ_2 and ρ_2 are constants (4.5, 4) that represent the decay rate of the membrane potential and the strength of the inhibition, respectively.

V1 layer 2/3 monocular complex cells. V1 layer 2/3 consists of both monocular and binocular complex cells. V1 layer 2/3 monocular complex cells pool output signals from layer 3B monocular simple cells of like orientation and both contrast polarities at each position, and receive surface-to-boundary feedback from the ipsilateral V1 FIDO. This feedback is a new model mechanism. At steady-state, their membrane potentials, $C_{ijk}^{L/R,s/l}$, are given by:

$$C_{ijk}^{L/R,s/l} = \left([B_{ijk}^{L/R,s/l,+}]^+ + [B_{ijk}^{L/R,s/l,-}]^+\right)\left(1 + f_1 \sum_d F_{(i\mp e)jkd}^{L/R,s/l,1}\right), \quad (24)$$

where $B_{ijk}^{L/R,s/l,+/-}$ are signals from V1 layer 3B monocular simple cells defined by equation (20), f_1 is a constant (0.5) that represents the relative strength of the feedback connections, $F_{ijkd}^{L/R,s/l,1}$ are contrast-selective surface-to-boundary feedback signals from V1 blobs that are defined by equation (31).

The feedback signals selectively enhance activities of the recipient V1 layer 2/3 complex cells. For V1 layer 2/3 monocular complex cells, feedback signals from the left FIDO (indicated by the upper index *L* of feedback signal *F*) of V1 blobs only enhance the activities of the left monocular boundary cells (indicated by the upper index *L* of complex cell activity *C*) but not the right monocular cells, and the allelotropic shift function is *(i-e)* in this case (indicated by the first lower index of feedback signal *F*). Similarly, feedback signals from the right FIDO (indicated by the upper index *R* of feedback signal *F*) of V1 blobs only enhance the activities of the right monocular boundary cells (indicated by the upper index *R* of complex cell activity *C*) but not the left monocular cells, and the allelotropic shift function is *(i+e)*. The effects of the feedback are modulatory, and can only enhance those boundaries that already have bottom-up inputs from lower layers, but cannot evoke boundaries at positions that have no bottom-up inputs.

V1 layer 2/3 binocular complex cells. V1 layer 2/3 binocular complex cells pool the cell membrane potentials of layer 3B binocular simple cells of like orientation and both contrast polarities (+/-) at each position, and receive surface-to-boundary feedback from both left and right monocular FIDOs in V1 blobs. This feedback is also a new model mechanism. At steady-state, their membrane potentials, $C_{ijk}^{B,s/l}$, are given by:

$$C_{ijkd}^{B,s/l} = \left([B_{ijkd}^{B,s/l,+}]^+ + [B_{ijkd}^{B,s/l,-}]^+\right)\left[1 + f_1\left(F_{ijkd}^{L,s/l,1} + F_{ijkd}^{R,s/l,1}\right)\right], \quad (25)$$

where $B_{ijkd}^{B,s/l,+/-}$ are signals from V1 layer 3B binocular simple cells defined by equation (21), f_1 is a constant (0.5) that represents the relative strength of the feedback connections, and $F_{ijkd}^{L/R,s/l,1}$ are surface-to-boundary feedback signals from V1 blobs that are defined by equation (31). Feedback signals from both the left and right FIDOs of V1 blobs selectively enhance activities of V1 layer 2/3 binocular complex cells. The effects of feedback are again modulatory.

V1 blobs. V1 blobs receive surface-feature signals from the LGN and boundary signals from the layer 2/3 binocular complex cells in the V1 interblobs. Left/right and ON/OFF monocular FIDOs fill-in with surface-feature signals from the ON-channel/OFF-channel of left/right LGNs, respectively. Left and right monocular FIDOs are separated from each other, while ON-channel and OFF-channels undergo opponent interactions. The surface-feature signals fill-in by a nearest-neighbor diffusion process whose lateral spread is gated, and thereby contained, by boundary signals, following Grossberg and Todorović (1988). Activities of V1 left FIDO cells $z_{ijd}^{L,on/off,1}$ and right FIDO cells $z_{ijd}^{R,on/off,1}$ obey the boundary-gated diffusion equations:

$$\varphi_z \frac{d}{dt} z_{ijd}^{L,on/off,1} = -z_{ijd}^{L,on/off,1} + \sum_{(p,q)\in n(i,j)} (z_{pqd}^{L,on/off,1} - z_{ijd}^{L,on/off,1}) r_{pqijd}^{1} + X_{(i+e)j}^{L,on/off} \tag{26}$$

and

$$\varphi_z \frac{d}{dt} z_{ijd}^{R,on/off,1} = -z_{ijd}^{R,on/off,1} + \sum_{(p,q)\in n(i,j)} (z_{pqd}^{R,on/off,1} - z_{ijd}^{R,on/off,1}) r_{pqijd}^{1} + X_{(i-e)j}^{R,on/off}, \tag{27}$$

where $X_{ij}^{L/R,on/off}$ are the surface-feature signals of Retina/LGN defined by equations (16) and (17). The boundary-gated diffusion coefficients r_{pqijd}^{1} obey:

$$r_{pqijd}^{1} = \frac{\mu}{1+\nu \sum_{(l,m)\in\omega(p,q)} \sum_{k} (C_{lmkd}^{B,s} + C_{lmkd}^{B,l})}, \tag{28}$$

where $C_{ijkd}^{B,s/l}$ are the activities of V1 layer 2/3 binocular complex cells, defined by equation (25). Both small-scale and the large-scale binocular complex cells in V1 layer 2/3 gate filling-in with equal strength.

Opponent ON-channel and OFF-channel inhibition at each position yields the double-opponent surface outputs $Z_{ijd}^{L/R,on/off,1}$:

$$Z_{ijd}^{L/R,on,1} = \left[z_{ijd}^{L/R,on,1} - z_{ijd}^{L/R,off,1} \right]^{+} \tag{29}$$

and

$$Z_{ijd}^{L/R,off,1} = \left[z_{ijd}^{L/R,off,1} - z_{ijd}^{L/R,on,1} \right]^{+} \tag{30}$$

Outputs of the contrast-sensitive cells, $F_{ijkd}^{L/R,on/off,1}$, are selective to the contrast between the successfully filled-in regions:

$$F_{ijkd}^{L/R,s/l,1} = \left[\left(\left[\sum_{p,q} b_{pqk}^{s/l} Z_{i+p,j+q,d}^{L/R,on,1} \right]^{+} + \left[\sum_{p,q} -b_{pqk}^{s/l} Z_{i+p,j+q,d}^{L/R,off,1} \right]^{+} + \left[\sum_{p,q} -b_{pqk}^{s/l} Z_{i+p,j+q,d}^{L/R,on,1} \right]^{+} + \left[\sum_{p,q} b_{pqk}^{s/l} Z_{i+p,j+q,d}^{L/R,off,1} \right]^{+} \right) - \theta_f^{s/l} \right]^{+} \tag{31}$$

The parameters and kernels are defined in Section A3.

V2 layer 4. In V2, virtually all cells are binocularly driven (Hubel and Livingstone, 1987), consistent with the model hypothesis that left and right monocular inputs are combined in V2 layer 4. 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.

Bottom-up inputs from V1 excite V2 layer 4 cells. V2 layer 4 cells also receive modulation from the surface-to-boundary feedback from V2 thin stripes. V2 layer 4 cells are enhanced by surface-to-boundary feedback at the same depth, and inhibited by surface-to-

boundary feedback at nearer depths and the same positions. At steady-state, their outputs, $J_{ijkd}^{s/l}$, are given by:

$$J_{ijkd}^{s/l} = \left[\Gamma_{ijkd}^{s/l} \left\{ 1 + f_2 \left[\left(F_{ijkd}^{L,s/l,2} + F_{ijkd}^{R,s/l,2} \right) - f_3 \sum_{d'<d} \left(F_{ijkd'}^{L,s/l,2} + F_{ijkd'}^{R,s/l,2} \right) \right] \right\} - \theta_J^{s/l} \right]^+ , \qquad (32)$$

where $\Gamma_{ijkd}^{s/l}$ are bottom-up inputs from V1 interblobs that combine the outputs of monocular and binocular complex cells:

$$\Gamma_{ijkd}^{s/l} = C_{ijkd}^{B,s/l} + \lambda \left(C_{(i+e)jk}^{L,s/l} + C_{(i-e)jk}^{R,s/l} \right) . \qquad (33)$$

In (33), λ is a constant (0.21) that represents the strength of the monocular connections, and $C_{ijkd}^{L/R/B,s/l}$ are the boundary signals from monocular or binocular complex cells in layer 2/3 of V1 interblobs, which are defined by equations (24) and (25). In addition, f_2 is a constant (0.5) that represents the relative strength of the feedback connections; f_3 is a constant (3) that represents the relative strength of the near-to-far inhibition in the surface-to-boundary feedback; (θ_J^s, θ_J^l) are constants (0.5, 1.5) that represent the thresholds at either the small scale or the large scale; and $F_{ijkd}^{L/R,s/l,2}$ are contrast-selective surface-to-boundary contrast-sensitive feedback signals from V2 thin stripes, which are defined in equation (50).

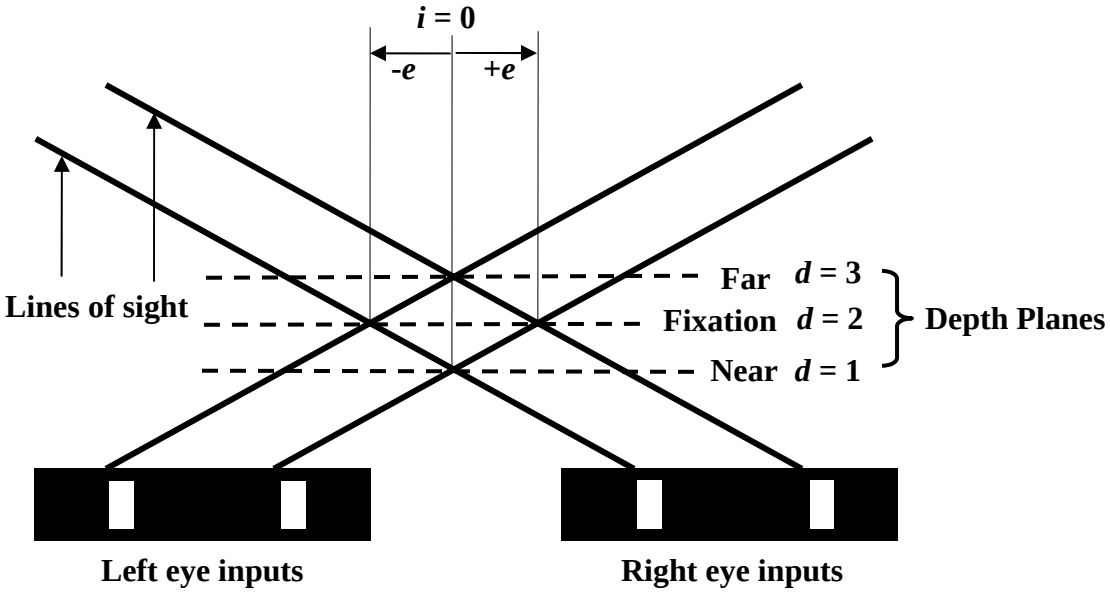


Figure 9. The V2 disparity filter: the V1 binocular boundaries network matches an edge in one retinal image with every other edge in the other retinal image whose relative disparity is not too great, that has the same contrast polarity and whose magnitude of contrast is not too different. In response to this image, the V1 boundary network creates four matches, with the two matches not in the fixation plane being

false matches between edges that do not correspond to the same object. These false matches are suppressed by the disparity filter in V2. Here, each neuron is inhibited by every other neuron that shares either of its monocular inputs; that is, shares a monocular line-of-sight represented by the solid lines. The solid lines also help to define the allelotropic shifts (i.e., *e* in the figure): an edge in the left (right) retinal image is shifted to the right (left) for matches increasingly further away.

V2 layer 2/3 bipole cells. Pyramidal cells in V2 layer 2/3 receive inputs from V2 layer 4 cells. These 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 pyramidal cells as well as nearby inhibitory interneurons. These pyramidal cells in V2 layer 2/3 are called *bipole cells* because the balance of excitation and inhibition helps to implement a bipole property that controls boundary grouping (Grossberg, 1994; Grossberg and Mingolla, 1985; Grossberg and Raizada, 2000; von der Heydt, Peterhans, and Baumgartner, 1984). The model also proposes that three other types of inhibition are all part of the process of 3D perceptual grouping: the disparity filter, in which cells that represent the same line-of-sight inhibit each other (Figure 9); the spatial competition between nearby cells that are tuned to the same orientation; and the orientation competition between nearby cells that are tuned to different orientations. Accordingly, the membrane potential, $T_{ijkd}^{s/l}$, of the bipole cell in V2 layer 2/3 at position *(i,j)* that codes orientation *k*, disparity *d* and either small (*s*) or large (*l*) scale, obeys the equation:

$$\frac{d}{dt}T_{ijkd}^{s/l} = -\varepsilon T_{ijkd}^{s/l} + (\alpha - T_{ijkd}^{s/l})\left[\left[J_{ijkd}^{s/l}\right]^{+} + H_{ijkd}^{s/l,u} + H_{ijkd}^{s/l,v}\right]^{+} - T_{ijkd}^{s/l}\left[\eta_1\left(\left[P_{ijkd}^{s/l,u}\right]^{+} + \left[P_{ijkd}^{s/l,v}\right]^{+}\right) + \Omega_{ijd}^{H/V,s/l} + \Lambda_{ijd}^{H/V,s/l}\right], \quad (34)$$

where $J_{ijd}^{H/V,s/l}$ are the activities of the V2 layer 4 cells defined by equation (32); and η_1 is a constant (3) that scales the inhibition from the inhibitory cells. Terms $H_{ijkd}^{s/l,u}$ and $H_{ijkd}^{s/l,v}$ describe excitatory inputs from the long-range connections in V2 layer 2/3 to a bipole cell at position (*i,j*), orientation *k*, disparity *d*, and scale *s* or *l*. Index *u/v* means that the input comes from side *u* or side *v* of the bipole cell. These terms obey:

$$H_{ijkd}^{s/l,u} = \sum_{pq} h_{pqijk}^{s/l,u}\left[T_{ijkd}^{s/l} - \theta_t\right]^{+} \quad (35)$$

and

$$H_{ijkd}^{s/l,v} = \sum_{pq} h_{pqijk}^{s/l,v}\left[T_{ijkd}^{s/l} - \theta_t\right]^{+}, \quad (36)$$

where θ_t is a constant signal threshold (3). Kernels $h_{pqijk}^{s/l,u}$ and $h_{pqijk}^{s/l,v}$ are the long-range connection weights in the scale *s* or *l*, at orientation *k*, and side *u* or *v* from the bipole cell at position (*p,q*) to the bipole cell at position (*i,j*). The connection weights when *k* refers to the horizontal orientation are defined as follows:

$$h_{pqijk}^{s/l,u} = sign(i-p)\phi_h^{s/l}\exp\left[-\frac{\eta_h(p-i)^2 + (q-j)^2}{\left(\delta_h^{s/l}\right)^2}\right] \quad (37)$$

and

$$h_{pqijk}^{s/l,v} = sign(p-i)\phi_h^{s/l} \exp\left[-\frac{(p-i)^2 + \eta_h (q-j)^2}{\left(\delta_h^{s/l}\right)^2}\right], \tag{38}$$

where $sign(x) = 1$ if $x > 0$, and 0 otherwise. Parameters $(\eta_k, \phi_h^s, \delta_h^s, \phi_h^l, \delta_h^l)$ are constants (25, 0.5, 7, 0.25, 70). The connection weights when *k* refers to the vertical orientation are obtained by a rotation of 90 degrees.

Terms $P_{ijkd}^{s/l,u}$ and $P_{ijkd}^{s/l,v}$ in equation (34) represent activities of the inhibitory interneurons associated with the bipole cell at position *(i,j)*, orientation *k*, disparity *d*, scale *s* or *l*, and side *u* or *v*. There is an inhibitory interneuron associated with each side (*u* or *v*) of a bipole cell. This inhibition from the inhibitory interneurons helps to maintain the balance between excitation and inhibition to enforce the bipole property. Activities $P_{ijkd}^{s/l,u}$ and $P_{ijkd}^{s/l,v}$ are defined by:

$$\varphi_p \frac{dP_{ijkd}^{s/l,u}}{dt} = -P_{ijkd}^{s/l,u} + H_{ijkd}^{s/l,u} - \beta_p P_{ijkd}^{s/l,u} [P_{ijkd}^{s/l,v}]^+ \tag{39}$$

and

$$\varphi_p \frac{dP_{ijkd}^{s/l,v}}{dt} = -P_{ijkd}^{s/l,v} + H_{ijkd}^{s/l,v} - \beta_p P_{ijkd}^{s/l,v} [P_{ijkd}^{s/l,u}]^+, \tag{40}$$

where positive rate parameter $\varphi_p << 1$ represents the rapid responses of the inhibitory cells; $H_{ijkd}^{s/l,u}$ and $H_{ijkd}^{s/l,v}$ are excitatory inputs on its side (*u* or *v*) from V2 layer 2/3 pyramidal cells, defined by equations (35) and (36); and β_p is a constant (1) that scales the recurrent inhibition from the other group of inhibitory cells at the same location. Equations (34), (39), and (40) show that inhibitory interneurons inhibit each other as well as their target pyramidal cells. Solving both the equations for $P_{ijkd}^{s/l,u}$ and $P_{ijkd}^{s/l,v}$ together at steady state, one finds:

$$P_{ijkd}^{s/l,u} = \frac{1}{2\beta_p}(-B_u + \sqrt{B_u^{\,2} + 4\beta_p H_{ijkd}^{s/l,u}}) \tag{41}$$

and

$$P_{ijkd}^{s/l,v} = \frac{1}{2\beta_p}(-B_v + \sqrt{B_v^{\,2} + 4\beta_p H_{ijkd}^{s/l,v}}), \tag{42}$$

where $B_u = 1 + \beta_p (H_{ijkd}^{s/l,v} - H_{ijkd}^{s/l,u})$ and $B_v = 1 + \beta_p (H_{ijkd}^{s/l,u} - H_{ijkd}^{s/l,v})$. These steady-state values are used in the simulations.

A bipole cell has two oriented branches (e.g., left and right for the horizontal orientation), each of which is associated with one inhibitory interneuron (population). Therefore, there are two inhibitory interneurons associated with each bipole cell. In equations (41) and (42), only the related two inhibitory cells inhibit each other. A bipole cell will not fire when it receives excitatory input from only one side of its long-range connection, but it can fire when it receives excitatory inputs from both sides. In equations (39) and (40), for example, when $H_{ijkd}^{s/l,u}$ equals zero, then $P_{ijkd}^{s/l,u}$ equals zero. As a result, $P_{ijkd}^{s/l,v}$ equals $H_{ijkd}^{s/l,v}$. Therefore, the total excitatory input from the long-range connections equal the total inhibitory input from the inhibitory interneurons, and cannot fire the target bipole cell. But when both $H_{ijkd}^{s/l,u}$ and $H_{ijkd}^{s/l,v}$ are far from zero, the total excitatory inputs are much larger than the inhibitory inputs, because of the mutual

interneuron inhibition, and thus the target bipole cell can fire. In particular, a collinear bipole cell in V2 layer 2/3 which receives large excitatory inputs from both sides via its long-range horizontal intra-layer connections can fire, even when it receives no direct input from V2 layer 4.

Term $\Omega_{ijkd}^{s/l}$ in equation (34) is the total line-of-sight inhibition that a bipole cell receives from all other depth planes. This term is the disparity filter:

$$\Omega_{ijkd}^{s/l} = \eta_2 \sum_{d' \neq d} M_{dd'} ([T_{(i+e'-e)jkd'}^{s/l} - \theta_t]^+ + [T_{(i+e-e')jkd'}^{s/l} - \theta_t]^+), \tag{43}$$

where η_2 is a constant (50) that scales the total inhibition that each cell receives from the disparity filter; and the coefficient $M_{dd'}$ (listed in Table 2) are the strengths of line-of-sight inhibition from all other cells that share a monocular input between disparities *d* and *d'*.

	Near	**Fixation**	**Far**
Near	-	2.5	2
Fixation	1.5	-	1.5
Far	2	2.5	-

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 9).

Term $\Lambda_{ijkd}^{s/l}$ in equation (34) describes the spatial and orientational competition that a bipole cell receives from nearby cells:

Term $\Lambda_{ijkd}^{s/l}$ in equation (34) describes the spatial and orientational competition that a bipole cell receives from nearby cells:

$$\Lambda_{ijkd}^{s/l} = \eta_3 \sum_{pq} g_{pqij}^{s/l} \left([T_{pqkd}^{s/l} - \theta_t]^+ + \eta_4 \left[\sum_{k' \neq k} \left(T_{pqk'd}^{s/l} - \theta_t \right) \right]^+ \right), \tag{44}$$

where η_3 is a constant (200) that scales the total inhibition that each cell receives from the orientational competition; $g_{pqij}^{s/l}$ are 2D Gaussian kernels that are defined in equation (14), except that the kernel parameters (ϕ_g, σ_g) are (1.25, 0.8) for the small-scale kernel g_{pqij}^{s}, and (4, 6.4) for the large-scale kernel g_{pqij}^{l}, respectively; and η_4 is a constant (0.5) that represents the relative strength of the orientational competition compared to the spatial competition.

V2 thin stripes. V2 thin stripes receive surface-feature signals from the LGN and boundary signals from the layer 2/3 complex cells in the V2 pale stripes. Left/right and ON/OFF monocular FIDOs fill-in with surface-feature signals from the ON-channel/OFF-channel of left/right LGNs, respectively. Left and right monocular FIDOs are separated from each other in V2, as in V1. ON-channel and OFF-channel again form mutually inhibitory opponent channels. Activities of V2 left FIDO cells $z_{ijd}^{L,on/off,2}$ and right FIDO cells $z_{ijd}^{R,on/off,2}$ obey the boundary-gated diffusion equations:

$$\varphi_z \frac{d}{dt} z_{ijd}^{L,on/off,2} = -z_{ijd}^{L,on/off,2} + \sum_{(p,q)\in n(i,j)} (z_{pqd}^{L,on/off,2} - z_{ijd}^{L,on/off,2}) r_{pqijd}^2 + X_{(i+e)j}^{L,on/off} \quad (45)$$

and

$$\varphi_z \frac{d}{dt} z_{ijd}^{R,on/off,2} = -z_{ijd}^{R,on/off,2} + \sum_{(p,q)\in n(i,j)} (z_{pqd}^{R,on/off,2} - z_{ijd}^{R,on/off,2}) r_{pqijd}^2 + X_{(i-e)j}^{R,on/off} \ , \quad (46)$$

where $X_{ij}^{L/R,on/off}$ are the surface-feature signals of Retina/LGN defined by equations (16) and (17). The boundary-gated diffusion coefficients r_{pqijd}^2 obey:

$$r_{pqijd}^2 = \frac{\mu}{1 + \nu \sum_{(l,m)\in\omega(p,q)} \sum_k ([T_{lmkd}^s - \theta_t]^+ + [T_{lmkd}^l - \theta_t]^+)}, \quad (47)$$

where $T_{ijkd}^{s/l}$ are the activities of V2 layer 2/3 complex cells in equation (34). Opponent ON-channel and OFF-channel inhibition at each position yields the surface outputs $Z_{ijd}^{L/R,on/off,2}$:

$$Z_{ijd}^{L/R,on,2} = [z_{ijd}^{L/R,on,2} - z_{ijd}^{L/R,off,2}]^+ \quad (48)$$

and

$$Z_{ijd}^{L/R,off,2} = [z_{ijd}^{L/R,off,2} - z_{ijd}^{L/R,on,2}]^+ \quad (49)$$

Outputs of the contrast-sensitive cells, $F_{ijkd}^{L/R,on/off,2}$, are selective to the contrast between the successfully filled-in regions:

$$F_{ijkd}^{L/R,s/l,2} = \left[\left(\left[\sum_{p,q} b_{pqk}^{s/l} Z_{i+p,j+q,d}^{L/R,on,2} \right]^+ + \left[\sum_{p,q} -b_{pqk}^{s/l} Z_{i+p,j+q,d}^{L/R,off,2} \right]^+ + \left[\sum_{p,q} -b_{pqk}^{s/l} Z_{i+p,j+q,d}^{L/R,on,2} \right]^+ + \left[\sum_{p,q} b_{pqk}^{s/l} Z_{i+p,j+q,d}^{L/R,off,2} \right]^+ \right) - \theta_f^{s/l} \right]^+ \quad (50)$$

The parameters and kernels are defined in Section A3.

V4. V4 gives rise to visible 3D surface percepts. V4 cells receive surface-feature signals from the LGN via the V1 blobs and the V2 thin stripes, and boundary signals from the V2 pale stripes. V4 forms binocular surface representations that combine features from both eyes. Furthermore, nearer boundaries are added to farther depths in position (boundary enrichment), and successfully filled-in features in V2 thin stripes are subtracted from farther depths (surface pruning) in V4 to ensure that opaque objects can be seen. Activities of V4 binocular FIDO cells $z_{ijd}^{B,on/off,4}$ obey the boundary-gated diffusion equation:

$$\varphi_z \frac{d}{dt} z_{ijd}^{B,on/off,4} = -z_{ijd}^{B,on/off,4} + \sum_{(p,q)\in n(i,j)} (z_{pqd}^{R,on/off,4} - z_{ijd}^{B,on/off,4}) r_{pqijd}^4 + y_{ijd}^{on/off} \ , \quad (51)$$

where $y_{ijd}^{on/off}$ are the surface-feature signals that have been pruned by the successfully filled-in surfaces at the nearer depths:

$$y_{ijd}^{on/off} = \left[X_{(i+e)j}^{L,on/off} + X_{(i-e)j}^{R,on/off} - \tau_1 \sum_{d'<d} \sum_k \left(F_{(i+e-e')jkd'}^{L,s,2} + F_{(i-e+e')jkd'}^{L,s,2} + F_{(i+e-e')jkd'}^{R,s,2} + F_{(i-e+e')jkd'}^{R,s,2} \right) \right]^+ , \quad (52)$$

where $X_{ij}^{L/R,on/off}$ are the surface-feature signals of the Retina/LGN defined by equations (16) and (17); τ_1 is a constant (3) that represents the pruning strength; and $F_{ijkd}^{L/R,s,2}$ are the contrast-sensitive outputs from model V2 thin stripes and are defined by equation (50).

In equation (51), r^4_{pqijd} are the boundary-gated diffusion coefficients that are determined by boundary signals:

$$r^4_{pqijd} = \frac{\mu}{1+\nu \sum\limits_{(l,m)\in\omega(p,q)} \sum\limits_{k} (a^s_{lmkd} + a^l_{lmkd})}, \tag{53}$$

where $a^{s/l}_{lmkd}$ are boundary signals from V2 pale stripes that are derived from a boundary enrichment process:

$$a^{s/l}_{ijkd} = \left[T^{s/l}_{ijkd} - \theta_t\right]^+ + \tau_2 \sum_{d'<d} \left(\left[T^{s/l}_{(i+e-e')jkd'} - \theta_t\right]^+ + \left[T^{s/l}_{(i-e+e')jkd'} - \theta_t\right]^+ \right), \tag{54}$$

where $T^{s/l}_{ijkd}$ are the activities of layer 2/3 complex cells in the V2 pale stripes, defined by equation (34), and τ_2 is a constant (1) that represents the strength of boundary enrichment.

Opponent ON-channel and OFF-channel inhibition at each position yields the surface outputs $Z^{B,on/off,4}_{ijd}$:

$$Z^{B,on,4}_{ijd} = \left[z^{B,on,4}_{ijd} - z^{B,off,4}_{ijd}\right]^+ \tag{55}$$

and

$$Z^{B,off,4}_{ijd} = \left[z^{B,off,4}_{ijd} - z^{B,on,4}_{ijd}\right]^+ \tag{56}$$

References

Andrews, B. W., & Pollen, D. A. (1979). Relationship between spatial frequency selectivity and receptive field profile of simple cells. *Journal of Physiology (London), 287,* 163-176.

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

Borg-Graham, L. J., Monier, C., & Fregnac, Y. (1998). Visual input evokes transient and strong shunting inhibition in visual cortical neurons. *Nature, 393,* 369–373.

Cao, Y. & Grossberg, S. (2005). A laminar cortical model of stereopsis and 3D surface perception: closure and da Vinci stereopsis. *Spatial Vision, 18(5),* 515-578.

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

Cumming, B. G., & Parker, A. J. (2000). Local disparity not perceived depth is signaled by binocular neurons in cortical area V1 of the macaque. *Journal of Neuroscience, 20,* 4758-4767.

Dev, P. (1975). Perception of depth surfaces in random-dot stereograms: A neural model. *Int. J. Man-Machine Stud, 7*, 511-528.

DeYoe, E.A., & van Essen, D.C. (1988). Concurrent processing streams in monkey visual cortex. *Trends in Neurosciences*, *11*: 219-226.

Egusa, H. (1983). Effects of brightness, hue, and saturation on perceived depth between adjacent regions in the visual field. *Perception,* 12:167-175.

Faubert, J., and von Grunau, M. (1995) The influence of two spatially distinct primers and attribute priming on motion induction. *Vision Research, 35*, 3119–3130.

Fazl, A., Grossberg, S., & Mingolla, E. (2007). View-invariant object category learning, recognition, and search: How spatial and object attention are coordinated using surface-based attentional shrouds. Technical Report CAS/CNS-2007-011, Boston University. Submitted for publication.

Fleet, D. J., Wagner, H., & Heeger, D. J. (1996). Neural encoding of binocular disparity: Energy models, position shifts and phase shifts. *Vision Research, 36*, 1839-1857.

Foster, K. H., Gaska, J. P., Nagler, M., & Pollen, D. A. (1985). Spatial and temporal frequency selectivity of neurons in visual cortical areas V1 and V2 of the macaque monkey. *Journal of Physiology (London), 365,* 331–363.

Gilchrist, A. L. (1977). Perceived lightness depends on perceived spatial arrangement. *Science,* 195(4274):185-187.

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

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. (1973). Contour enhancement, short-term memory, and constancies in reverberating neural networks. *Studies in Applied Mathematics, 52,* 213-257.

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

Grossberg, S. (1987). Cortical dynamics of three-dimensional form, color, and brightness perception. *Perception and Psychophysics, 41,* 87-158.

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

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

Grossberg, S., & Hong, S. (2006). A neural model of surface perception: Lightness, anchoring, and filling-in. *Spatial Vision, 19(2-4),* 263-321(59).

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

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

Grossberg, S., & Raizada, R. (2000). Contrast-sensitive perceptual grouping and object-based attention in the laminar circuits of primary visual cortex. *Vision Research, 40,* 1413-1432.

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

Helmholtz, H. L. F. (1866). *Treatise on Physiological Optics*. Leopold Voss, Leipzig, Germany, reprinted by The Optical Society of America, Rochester, NY, 1924-1925.

Howard, I. P., & Rogers, B. J. (2002). *Seeing in Depth.* Toronto: I. Porteous.

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

Hubel, D. H., & Livingstone, M. S. (1987). Segregation of form, color, and stereopsis in primate area 18. *Journal of Neuroscience, 7(11),* 3378-3415.

Julesz, B. (1971). *Foundations of cyclopean perception.* University of Chicago Press, Chicago.

Kanizsa, G. (1974). Contours without gradients or cognitive contours. *Italian Journal of Psychology, 1*, 93–113.

Land, E., & McCann, J. J. (1971). Lightness and retinex theory. *Journal of the Optical Society of America, 61,* 1-11.

Levitt, J. B., Yoshioka, T. & Lund, J. S. (1994). Intrinsic cortical connections in macaque visual area V2: evidence for interaction between different functional streams. *Journal of Comparative Neurology, 342,* 551–570.

Malach, R. (1992). Dendritic sampling across processing streams in monkey striate cortex. *Journal of Comparative Neurology, 315,* 303 – 312.

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

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

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

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

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

Qian, N. (1994). Computing stereo disparity and motion with known binocular cell properties. *Neural Computation, 6,* 390–404.

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

Pessoa, L., Beck, J. and Mingolla, E. (1996). Perceived texture segregation in chromatic element-arrangement patterns: High intensity interference. *Vision Research, 36,* 1745–1760.

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

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

Rockland, K. S., & Lund, J. S. (1983). Intrinsic laminar lattice connections in primate visual cortex. *Journal of Comparative Neurology, 216,* 303–318.

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.

Sperling, G. (1970). Binocular vision: a physical and a neural theory. *American Journal of Psychology, 83,* 461-534.

Tootell, R. B. H., Silverman, M. S., De Valois, R. L. (1981). Spatial frequency columns in primary visual cortex. *Science, 214,* 813-815.

Ts'o, D. Y., Roe, A. W., & Gilbert, C. D. (2001). A hierarchy of the functional organization for color, form and disparity in primate visual area V2. *Vision Research, 41(10-11),* 1333-1349.

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

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.