Pergamon

Neural Networks, Vol. 10, No. 9, pp. 1583–1605, 1997

Printed in Great Britain
0893–6080/97 $17.00+.00

PII: S0893-6080(97)00065-8

CONTRIBUTED ARTICLE

Cortical Dynamics of Three-Dimensional Surface Perception: Binocular and Half-Occluded Scenic Images

STEPHEN GROSSBERG[1] AND NIALL P. MCLOUGHLIN[2]

[1]Department of Cognitive and Neural Systems and Center for Adaptive Systems, Boston University and
[2]Department of Neurobiology, Harvard Medical School

(*Received* 14 *January* 1997; *accepted* 30 *June* 1997)

Abstract—*Previous models of stereopsis have concentrated on the task of binocularly matching left and right eye primitives uniquely. A disparity smoothness constraint is often invoked to limit the number of possible matches. These approaches neglect the fact that surface discontinuities are both abundant in natural everyday scenes, and provide a useful cue for scene segmentation. da Vinci stereopsis refers to the more general problem of dealing with surface discontinuities and their associated unmatched monocular regions within binocular scenes. This study develops a mathematical realization of a neural network theory of biological vision, called FACADE theory, that shows how early cortical stereopsis processes are related to later cortical processes of three-dimensional surface representation. The mathematical model demonstrates through computer simulation how the visual cortex may generate three-dimensional boundary segmentations and use them to control filling-in of three-dimensional surface properties in response to visual scenes. Model mechanisms correctly match disparate binocular regions while filling-in monocular regions with the correct depth within a binocularly viewed scene. This achievement required the introduction of a new multiscale binocular filter for stereo matching which clarifies how cortical complex cells match image contours of like contrast polarity, while pooling signals from opposite contrast polarities. The filter also suggests how false binocular matches and unmatched monocular cues are automatically handled across multiple spatial scales. Pooling of signals from even- and odd-symmetric simple cells at complex cells helps to eliminate spurious activity peaks in matchable signals. Later stages of cortical processing by the blob and interblob streams, including refined models of cooperative boundary grouping and reciprocal stream interactions between boundary and surface representations, are modeled to provide a complete simulation of the da Vinci stereopsis percept.* © *1997 Elsevier Science Ltd. All rights reserved.*

Keywords—Binocular vision, Boundary contour system, da Vinci stereopsis, Depth perception, FACADE theory, Feature contour system, Neural network, Stereopsis, Visual cortex.

1. INTRODUCTION

Scientists since the time of Euclid (1557) have noted that "the pictures of bodies seen by both eyes are formed by the union of two dissimilar pictures formed by each" (Brewster, 1856). Until recently, most computational models of this binocular combination process have concentrated on solving the correspondence problem (Marr, 1980). This is the problem of binocularly matching identical regions from each eye's view into a single cyclopean percept. Typical approaches employ a variety of constraints to simplify this problem even further, one example being the uniqueness principle of Marr (1980). However, because the eyes are horizontally displaced within the head and therefore receive different two-dimensional images of the world, each eye typically sees regions that are not registered by the other eye. In pursuing the correspondence problem, many researchers have neglected the fact that these unmatched monocular regions are common. For example, when we view a partially occluded surface, one eye typically registers more of the farther surface than the other does. Our percept of the farther surface is derived from an integration of the

Acknowledgements: S.G. was supported in part by the Defense Advanced Research Projects Agency (ONR N00014-92-J-4015) and the Office of Naval Research (ONR N00014-95-1-0409 and ONR N00014-85-1-0657). N.P.M was supported in part by the Airforce Office of Scientific Research (AFOSR 90-0175), the National Science Foundation (NSF IRI-90-00530), the Defense Advanced Research Projects Agency (ONR N00014-92-J-4015), the Office of Naval Research (ONR N00014-91-J-4100), and The James S. McDonnell Foundation 94-40. The authors also wish to thank Rebekah Brown and Robin Locke for their valuable assistance in the preparation of the manuscript.

Requests for reprints should be sent to: Professor S. Grossberg, Department of Cognitive and Neural Systems and Center for Adaptive Systems, Boston University, 677 Beacon Street, Boston, MA 02215, USA; e-mail: steve@cns.bu.edu.

binocularly viewed region with the monocular representation seen by one of the eyes. These "half-occluded" (Belhumeur & Mumford, 1992; Anderson & Nakayama, 1994) regions are perceived at the correct depth and form an integrated part of the three-dimensional percept.

Nakayama and Shimojo (1990) coined the term da Vinci stereopsis to describe this phenomenon, although da Vinci was certainly not the first to describe it (Brewster, 1856). Galen unambiguously described this phenomenon over 1500 years ago when he suggested "Standing near a column, and shutting each of the eyes in succession—when the right eye is shut, some of those parts of the column which were previously seen by the right eye on the right side of the column, will not now be seen by the left eye; and when the left eye is shut, some of those parts which were formerly seen on the left side of the column, will not now be seen by the right eye. But when we, at the same time, open both eyes, both these will be seen, for a greater part is concealed when we look with either of the two eyes, than when we look with both at the same time" (Galen, 1550/1968).

A more recent example of one such stimulus is presented in Figure 1. Observers see the right eye view of surface BD in depth although the region BC that lies between the vertical lines B and C is registered monocularly by only the right eye. This apparently innocuous stimulus presents a serious challenge to models of three-dimensional surface perception to explain how the monocularly viewed region BC becomes attributed to the correct depth.

da Vinci stereopsis is closely related to the equidistance tendency, which has been studied extensively by Gogel (1956, 1965, 1970). Gogel noted that a monocularly viewed object in a binocular scene seems to lie at the same depth as the retinally most contiguous binocularly viewed object. Gogel used a complex mirror arrangement to ensure that one object in his visual scene was presented monocularly while all others were presented binocularly. Emmert (1881) had also reported a comparable percept. He discovered that a monocular afterimage appears to be located at whatever depth the subject binocularly fixates.

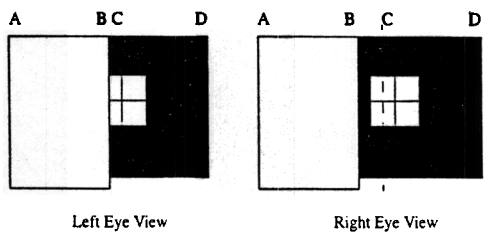

FIGURE 1. An example da Vinci stereogram derived from viewing a three-dimensional scene of a room. Regions AB and CD are presented with positive (crossed) disparity. Region BC is seen monocularly (in the right eye view only). This stereogram represents the retinal images impinging on an observer near an occlusion in depth caused by a nearer object (surface AB).

A second often overlooked aspect of the binocular combination process is allelotropia. Allelotropia is the phenomenon whereby a binocularly fused edge is formed by deforming two disparate monocular images into a single binocular one (Porta, 1593; Werner, 1937; von Tschermak-Seysenegg, 1952; Kaufman, 1974). It is more commonly referred to as displacement. A classical example of this phenomenon occurs when the pattern EF G is viewed through one eye and the pattern E FG is viewed through the other eye. The letter F is seen floating in depth at a location halfway between E and G. Neither the left nor the right image of F occupied this location. Thus the process of binocular fusion deforms the monocular representations of F into a single binocular percept of F whose spatial location differs from either of its monocular representations.

In a three-dimensional scene, objects at different depths, and hence different retinal disparities, require different amounts of displacement. Objects close to the fixation plane (zero disparity) require very little deformation while nonfoveated objects closer to the observer require much larger degrees of allelotropia. Hence different parts of the left and right eye views are deformed by different amounts to form a single binocular percept of the scene. In the da Vinci stereogram example (Figure 1), the vertical boundaries of the nearer wall (AB) are deformed to a larger degree than those of the back wall (CD). With all these deformations going on, one needs to explain how three-dimensional surface representations form in a perceptually seamless fashion.

da Vinci stereopsis percepts also raise issues concerning the role of occluding surfaces and eye-of-origin information in depth perception. Nakayama and Shimojo (1990) have developed da Vinci stereograms which show that both sorts of information are used to form the final percept. In particular, they described a stereogram in which unpaired monocular dots, without any binocular disparity, can generate a percept of a depthful occluding surface. The present model incorporates both monocular cells (and thus eye-of-origin information) and binocular cells in its stereo matching circuit. It also includes mechanisms of figure–ground separation that govern the formation of occluding surfaces in response to both three-dimensional scenes, which do include disparity information, and two-dimensional pictures, which do not.

More generally, the model describes how the earlier processes of stereopsis work and interact with later processes of three-dimensional boundary and surface perception. It presents an account of these interactions that includes a new multiscale binocular filter for stereo matching of both unmatched monocular and matched binocular stimuli. It shows how to design such a filter so that its inputs to the boundary system are appropriately designed to lead to effective three-dimensional boundary and surface percepts.

In particular, this filter clarifies why the brain prefers to binocularly match image contours with the same

contrast polarity (Anderson & Nakayama, 1994; von Helmholtz, 1910/1925), yet the complex cells at which they are matched pool signals from opposite contrast polarities (Hubel & Wiesel, 1965; Pollen et al., 1989). Pooling signals from opposite contrast polarities enables the brain to form object boundaries around objects whose relative contrasts with their backgrounds reverse along their perimeter (Grossberg, 1987a; Grossberg and Mingolla, 1985b). Thus the new multiscale binocular filter clarifies how the brain begins to reconcile constraints on binocular matching with constraints on the formation of depthful object boundaries. After these early monocular or binocular matches are made, competitive and cooperative interactions within the boundary system work to ensure that only the correct subset of matches survive to form the three-dimensional boundary and surface representations that are perceived. Reciprocal and hierarchical interactions between the boundary and surface processing streams further ensure that their final representations are mutually consistent, even though the operations that lead to these representations are computationally complementary. Along the way, these processes compensate for allelotropia.

The present article develops these themes by providing the first mathematically rigorous implementation of a neural network theory of three-dimensional biological vision and figure–ground separation, called FACADE theory, that was introduced in Grossberg (1993, 1994, 1997) and used there to qualitatively explain a large number of three-dimensional percepts, including da Vinci stereopsis. The results of our implementation were first reported in McLoughlin and Grossberg (1994). FACADE theory acquires its name from the multiplexed representations of Form-And-Color-And-DEpth that it produces in response to both monocular and binocular images. These FACADE representations, which are proposed to occur in prestriate cortical area V4, model how the visual cortex generates boundary and surface representations in response to visual scenes. A boundary contour system (BCS) generates boundary segmentations in response to visual cues from edges, textures, shading, and stereo information. A feature contour system (FCS) compensates for variable illumination conditions and fills-in surface representations that combine properties of brightness, color, depth and form. FACADE theory models how the BCS and FCS are organized and interact hierarchically and in parallel to generate three-dimensional percepts of the world.

A detailed introduction to the theory and its explanatory scope is given in Grossberg (1994). Here it is reviewed only enough to frame the new results. Although the mathematical realization described herein is potentially capable of simulating all the results outlined in Grossberg (1994, 1997), the percept of da Vinci stereopsis was selected as an explanatory target because it presents a formidable challenge to all theories of three-dimensional vision. To provide a complete simulation of da Vinci stereopsis, we needed to further develop the theory's multiple-scale binocular filter to deal with the problems of false matches and unmatched boundaries (Marr & Poggio, 1976, 1979; Marr, 1980; Frisby & Pollard, 1991). This binocular filter then turned out to also be competent to simulate recent results concerning the uniqueness principle of classical stereopsis (McKee et al., 1994), stereo suppression of monocular cues (McKee et al., 1990; McKee & Harrad, 1993), and the contrast-sensitive properties of dichoptic masking (McKee, 1993; McKee et al., 1994). Anderson and Nakayama (1994) have described related examples of contrast-sensitive binocular matching. These data, which have not yet been accounted for by other theories of stereopsis, are simulated in McLoughlin and Grossberg (1997). The present work hereby develops a previously unsuspected link between da Vinci stereopsis and data concerning the spatial range and contrast sensitivity of binocular matching. It also develops a framework for analyzing the data of Nakayama and Shimojo (1990) on da Vinci stereopsis, who noted the importance of eye-of-origin information in binocular vision. The monocular cells that are introduced herein carry this type of information.

In a similar vein, Harris and Parker (1995) suggested that bright and dark information were processed by independent neural mechanisms. They measured the statistical efficiency of stereo depth detection for subjects viewing binocular random dot stimuli composed of dots of either one or two polarities (half brighter and half darker than the background). Subjects were approximately twice as good at detecting stereo depth if both polarities were present. They went on to show that it was contrast polarity that was important by testing subjects with stimuli composed of two sets of dots of the same polarity but different contrasts. Efficiency fell back to the same level as when one set of dots was present. The binocular matching scheme employed herein can account for these results, since binocular matching as defined below is sensitive to direction-of-contrast. Like contrast dots match while opposite contrast dots do not. By changing the polarity of half the dots in the stimulus, less false matches are generated as opposite polarity dots are not matched. On the other hand, the cortical complex cells at which binocular matching occurs summate matches of either polarity. This also occurs in the model. Harris and Parker also comment upon such a possibility when they state that "although contrast polarity may be used to assist binocular matching in a population of disparity-selective neurons, the signal delivered by the output of such neurons may reflect only the disparity values and may fail to indicate which feature generated any particular disparity value" (Harris & Parker, 1995, p. 810).

2. FACADE THEORY MACROCIRCUIT

The processing stages of FACADE theory are summarized in Figure 2. Their functional role is briefly outlined

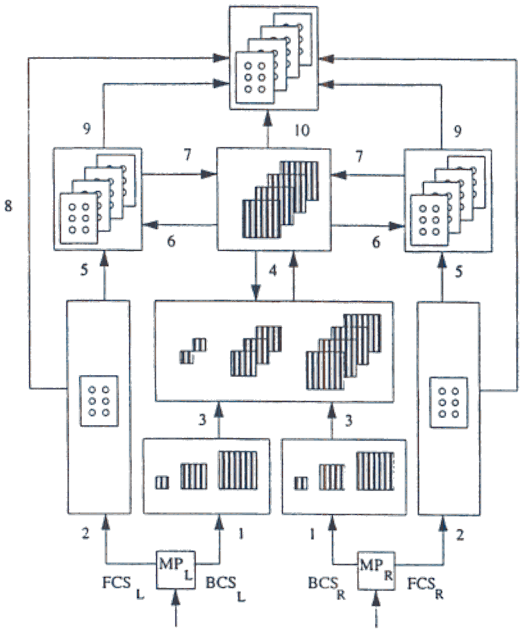

FIGURE 2. Macrocircuit of FACADE theory, showing monocular and binocular interactions of the BCS (boxes with vertical lines, designating oriented responses) and FCS (boxes with three pairs of circles, designating opponent colors). See text for details.

below. BCS processing stages are displayed as boxes with vertical lines that designate oriented responses. FCS stages are shown as boxes with three pairs of circles that denote opponent colors.

Monocular preprocessing of left eye (MP_L) and right eye (MP_R) inputs by retina and lateral geniculate nucleus (LGN) discounts the illuminant and generates parallel signals to BCS and FCS via pathways 1 and 2, respectively. Pathways 1 model monocular inputs to the interblobs in striate area V1. They activate model simple cells with oriented receptive fields that come in multiple sizes. Pathways 2 model monocular inputs to the blobs in striate area V1. They activate model blob cells that are tuned to opponent colors.

Pathways 3 support binocular combination of simple cell outputs at complex and complex end-stopped (or hypercomplex) cells. These interactions generate populations of disparity-sensitive cells that realize a size-disparity correlation. In particular, complex cells with larger receptive fields can binocularly fuse a broader range of disparities than can cells with smaller receptive fields. Competition across disparity at each position and among cells of a given size scale sharpens complex cell disparity tuning. Spatial competition (endstopping) and orientational competition convert complex cell responses into spatially and orientationally sharper responses at hypercomplex cells.

Pathways 4 initiate long-range horizontal grouping and boundary completion of the hypercomplex cell outputs by bipole cells. This grouping process collects together the outputs from all hypercomplex cells that are sensitive to a given depth range and inputs them to a shared set of bipole cells. The bipole cells, in turn, send cooperative feedback signals back to these hypercomplex cells. This feedback process binds together cells of multiple *sizes* into a BCS copy that is sensitive to a prescribed range of *depths*. In this way, each BCS copy completes boundaries within a given depth range. Multiple BCS copies are formed, each corresponding to different (but possibly overlapping) depth ranges.

These multiple depth-selective BCS copies are used to capture brightness and color signals within depth-selective FCS surface representations. These surface representations occur within monocular filling-in domains, or FIDOs, so called because they receive their brightness and color signals from a single eye, and support depth-selective filling-in of surface quality. A different monocular FIDO corresponds to each binocular BCS copy, although BCS copies that represent nearby depth ranges may send convergent signals, albeit with possibly different weights, to a single monocular FIDO.

Surface capture is achieved by a suitably defined interaction of BCS signals and illuminant-discounted FCS signals at the monocular FIDOs. The FCS signals reach the monocular FIDOs via pathways 5. These pathways carry out a one-to-many topographic registration of the monocular FCS signals at all the monocular FIDOs. Pathways 6 carry topographic BCS boundary signals from each BCS copy to its FIDO. These boundary signals selectively capture those FCS inputs from pathway 5 that are spatially coincident and orientationally aligned with the BCS boundaries. Other FCS inputs are suppressed by the BCS–FCS interaction.

The captured FCS inputs, and only these, can trigger diffusive filling-in of a surface representation on the corresponding FIDOs. Because this filled-in surface is activated by depth-selective BCS boundaries, it inherits the same depth as these boundaries. Not every triggered filling-in event can generate a surface representation. Because activity spreads until it hits a boundary, only surface regions that are surrounded by a connected BCS boundary, or fine web of such boundaries, are effectively filled-in. The diffusion of activity dissipates across the FIDO otherwise.

These BCS boundaries and FCS surfaces are formed by different, indeed complementary, processes. An analysis shows that too many boundary and surface fragments are formed as a result of the size-disparity correlation and of the way in which monocular and zero-disparity boundaries combine with nonzero disparity boundaries. Somehow these extra boundaries and surfaces need to be pruned. Pruning is realized by the process whereby the complementary boundary and surface properties interact to achieve boundary-surface consistency. Remarkably, many data about the perception of occluding and occluded objects may be explained as consequences of this pruning operation.

Boundary-surface consistency is achieved via pathways 7. Pathways 7 are activated by a contrast-sensitive

process that detects the contours of successfully filled-in surface regions at the monocular FIDOs. These FCS-to-BCS feedback signals excite the BCS boundaries corresponding to their own positions and depths. The boundaries that activated the successfully filled-in surfaces are hereby strengthened. The feedback signals also inhibit redundant boundaries at their own positions and larger depths. This inhibition from near-to-far is the first example within the theory of "the asymmetry between near and far". It is called boundary pruning. Boundary pruning spares the closest surface representation that successfully fills-in at a given set of positions.

Boundary pruning also removes redundant copies of the boundaries of occluding objects. When the competition from these redundant occluding boundaries is removed, the boundaries of partially occluded objects can be amodally completed behind them on BCS copies that represent larger depths. Moreover, when the redundant occluding boundaries collapse, the redundant surfaces that they momentarily supported at the monocular FIDOs collapse. Occluding surfaces are hereby seen to lie in front of occluded surfaces.

The surface representations that are generated at the monocular FIDOs are depth-selective, but they do not combine brightness and color signals from both eyes. Binocular combination of brightness and color signals takes place at the binocular FIDOs. Pathways 8 control the one-to-many topographic registration of the monocular FCS signals at all the binocular FIDOs, much as pathways 5 did for the monocular FIDOs. These FCS signals are binocularly matched at the binocular FIDOs. Only the surviving matched signals can be used for filling-in. These surviving matched signals are pruned by inhibitory signals from pathways 9. These inhibitory signals eliminate redundant FCS signals using the contrast-sensitive signals from the monocular FIDO surfaces that survive the boundary-surface consistency interactions of pathways 6 and 7. In particular, pathways 9 inhibit the FCS signals at their own positions and larger depths. As a result, occluding objects cannot redundantly fill-in surface representations at multiple depths. This is the second instance in the theory of the asymmetry between near and far. It is called surface pruning.

As in the case of the monocular FIDOs, the FCS signals to the binocular FIDOs can initiate filling-in only where they are spatially coincident and orientationally aligned with BCS boundaries. These boundaries are carried by pathways 10. These BCS-to-FCS pathways carry out depth-selective surface capture of the binocularly matched FCS signals from pathways 8 after they are pruned by inhibition from pathways 9.

The boundary signals along pathways 10 selectively capture those FCS signals that: (a) survive within-depth binocular FCS matching (pathways 8) and across-depth FCS inhibition (pathways 9); (b) are spatially coincident and orientationally aligned with the BCS boundaries; and (c) are surrounded by a connected boundary or fine web of such boundaries. Pathways 10 also realize the asymmetry between near and far through an operation that is called boundary enrichment. It adds the boundaries of near depths at the binocular FIDOs that represent larger depths. These additional boundaries prevent occluding objects from looking transparent by blocking filling-in of their occluded objects behind them. It can now be better seen how surface pruning and boundary enrichment work together: if boundary enrichment occurred without surface pruning, then the surfaces of occluding objects would be represented at all depths. If surface pruning occurred without boundary enrichment, then occluded objects could fill-in behind their occluders.

The total filled-in surface representation across all binocular FIDOs represents the visible percept. It is called a FACADE representation because it combines together, or multiplexes, properties of Form-And-Color-And-DEpth.

The model processing stages are neurophysiologically interpreted as follows (cf., de Yoe and van Essen, 1988). Monocular MP_L and MP_R preprocessing models those properties of retina and LGN that are required for present purposes. The BCS models the interblob cortical stream between cortical area V1 and V4, while the FCS models the blob stream. BCS simple, complex, hypercomplex and bipole processing is proposed to occur in the interblobs of V1 and the interstripes of V2. The monocular FIDOs are proposed to occur in V2 thin stripes, or possibly V1 blobs. The binocular FIDOs are proposed to occur in area V4. Bearing in mind that the BCS models the interblob cortical stream and the FCS the blob stream, the feedback signals between them clarify why the cells of these parallel cortical streams can be sensitive to shared combinations of features, despite their complementary functional roles.

3. MATHEMATICAL DESCRIPTION OF FACADE THEORY

The FACADE theory equations were simulated on a SUN Sparc workstation and presented with a binocular scene based on the da Vinci stereogram displayed in Figure 1. The processing stages are mathematically described in a step by step manner and their responses to the da Vinci stereopsis display are shown. These equations further develop the Boundary Contour System and Feature Contour System equations introduced in Grossberg and Mingolla (1985a, 1985b), Grossberg and Todorovic (1988), Gove et al. (1995) and Grossberg et al. (1995). As noted below, the simulated properties are robust, as many parameter choices yield similar results.

3.1. Input to the System

Input to the model consists of two views of the environment corresponding to the left and right eye inputs. These images for the da Vinci stereogram are displayed side by

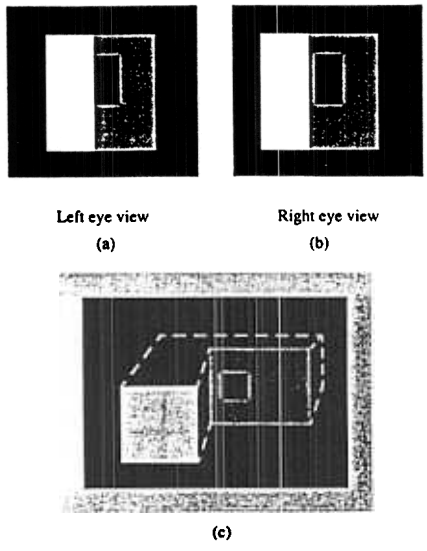

FIGURE 3. (a) Left and (b) right eye views of the da Vinci stereogram presented to the system. Note the monocularly viewed region containing the left-hand edge of the window and part of the far wall in the right eye view. (c) A schematic of the three-dimensional percept produced by fusing the da Vinci stereogram, as modeled in the text. Note that the side walls (dotted lines) are not perceived and are added only for clarification of the relative depths. The picture frame defines the plane of fixation (zero disparity).

side in Figure 3(a) and (b). The right eye view contains a region that is occluded in the left eye view. To be ecologically plausible, the occluded region must be relatively thin. The full three-dimensional layout is presented schematically in Figure 3(c). In the left eye's view, the near wall occludes part of the far wall. The model fixates the background which is therefore displayed.

4. CELL MEMBRANE EQUATIONS

Each model neuron was typically modeled as a single voltage compartment in which the membrane potential, $V(t)$, was given by

$$C_m \cdot \frac{dV(t)}{dt} = -(V(t) - E_{LEAK})g_{LEAK} - (V(t) - E_{EXCIT})g_{EXCIT}(t) - (V(t) - E_{INHIB})g_{INHIB}(t) \quad (1)$$

where the parameters E represent reversal potentials, g_{LEAK} is a constant leakage conductance, and the time-varying conductances $g_{EXCIT}(t)$ and $g_{INHIB}(t)$ represent the total inputs to the cell (Hodgkin, 1964; Grossberg, 1973). Transient afterhyperpolarization terms (AHP) were not incorporated since all groupings were allowed to reach steady state. The capacitance term C_m was set equal to 1 by rescaling time t. The leakage reversal potential E_{LEAK} was set equal to 0 by shifting the definition of $V(t)$. With this convention, the inhibitory reversal potential E_{INHIB} is nonpositive. Then eqn (1) can be rewritten in the form

$$\frac{d}{dt}V = -\alpha V + (U - V)g_{EXCIT} - (V + L)g_{INHIB} \quad (2)$$

where $\alpha = g_{INHIB}$ is a constant decay rate, $U = E_{EXCIT}$ and $L = |E_{INHIB}|$.

4.1. LGN ON and OFF Channels

The first processing level consists of a center-surround interaction corresponding to the ON and OFF channels of the retina and lateral geniculate body (Schiller, 1992). Within this implementation, the on-center off-surround (ON) cells, and the off-center on-surround (OFF) cells discount the illuminant and are sensitive to the ratio contrast of the local image region spanned by each cell's receptive field. The responses of the ON and OFF channels to the left and right da Vinci image views are half-wave rectified to generate outputs to both the BCS and the FCS.

Speaking mathematically, inputs I_{ij} are first processed by both ON cell activities x_{ij}^+ and OFF cell activities x_{ij}^- of cells centered on position ij. Each cell receives a Gaussian weighted sum of inputs (C) from a central region and an opponent Gaussian weighted sum of inputs (S) from the surround:

$$\frac{dx_{ij}^+}{dt} = -\alpha_1 x_{ij}^+ + (U_1 - x_{ij}^+)C_1 - (x_{ij}^+ + L_1)S_1 \quad (3)$$

$$\frac{dx_{ij}^-}{dt} = -\alpha_1 x_{ij}^- + (U_1 - x_{ij}^-)S_1 - (x_{ij}^- + L_1)C_1 \quad (4)$$

In eqns (3) and (4), the center and surround are defined by Gaussian kernels:

$$C_1 = \sum_{(p,q)} C_{pq} I_{i+p,j+q} \text{ and } S_1 = \sum_{(p,q)} S_{pq} I_{i+p,j+q} \quad (5)$$

such that

$$C_{pq} = A_1(2\pi\sigma_c^2)^{-1}\exp\left\{-\frac{1}{2}((p^2+q^2)/\sigma_c^2)\right\} \quad (6)$$

and

$$S_{pq} = A_2(2\pi\sigma_s^2)^{-1}\exp\left\{-\frac{1}{2}((p^2+q^2)/\sigma_s^2)\right\} \quad (7)$$

where $-P < p,q < P$. In addition, α_1 is a decay parameter (10); U_1 bounds the upper limit of cell activity (1); L_1 bounds the lower limit of cell activity (1); I_{ij} is the input at location ij; A_1 and A_2 are scale constants for the center and surround kernels (1, 1.03361); σ_c and σ_s are the standard deviations of the center and surround Gaussians (0.5, 1.5); and P defines the size of the center and surround (4).

At equilibrium, the ON and OFF activities converge to the self-normalizing equilibrium activities:

$$x_{ij}^+ = \frac{\sum_{(p,q)} U_1 C_{pq} I_{i+p,j+q} - L_1 S_{pq} I_{i+p,j+q}}{\alpha_1 + \sum_{(p,q)} (C_{pq} + S_{pq}) I_{i+p,j+q}} \quad (8)$$

and

$$x_{ij}^{-} = \frac{\sum_{(p,q)} U_1 S_{pq} I_{i+p,j+q} - L_1 C_{pq} I_{i+p,j+q}}{\alpha_1 + \sum_{(p,q)} (C_{pq} + S_{pq}) I_{i+p,j+q}} \quad (9)$$

The difference of these ON and OFF activities is computed to generate the LGN output signals (Grossberg, 1987b, 1994; Grossberg et al., 1995; Pessoa et al., 1995):

$$X_{ij}^{+} = [x_{ij}^{+} - x_{ij}^{-}]^{+}$$

$$= \frac{\left[(U_1 + L_1)\left(\sum_{(p,q)} C_{pq} - S_{pq}\right) I_{i+p,j+q}\right]^{+}}{\alpha_1 + \sum_{(p,q)} (C_{pq} + S_{pq}) I_{i+p,j+q}} \quad (10)$$

and

$$X_{ij}^{-} = [x_{ij}^{-} - x_{ij}^{+}]^{+}$$

$$= \frac{\left[(U_1 + L_1)\left(\sum_{(p,q)} S_{pq} - C_{pq}\right) I_{i+p,j+q}\right]^{+}}{\alpha_1 + \sum_{(p,q)} (C_{pq} + S_{pq}) I_{i+p,j+q}} \quad (11)$$

These signals depend on the Difference-of-Gaussians $C - S$, but are also sensitive to the net contrast $(C - S)(C + S)^{-1}$ of the local image region spanned by the kernels. This operation may be interpreted as occurring in either LGN or in V1.

4.2. Oriented Simple Cells

The first stage of the BCS (Figure 4) is composed of elongated contrast detectors that model simple cells in cortical area V1 (Hubel & Wiesel, 1965). Simple cells come in a variety of sizes, or spatial scales. In the current implementation, two spatial scales were implemented for each orientation. To reduce the complexity of this simulation, cells sensitive to either of two orientations were investigated: horizontal and vertical. Both the ON and the OFF channel information from LGN was processed by these simple cells.

Corresponding to each orientation and spatial scale are pairs of simple cells sensitive to two opposite contrast polarities: one for light/dark contrasts (S_{ijk}^{+}) and one for dark/light contrasts (S_{ijk}^{-}) Pooling of both contrast polarities occurs at the next level of BCS processing, the complex cells, by summing the half-wave rectified responses of both polarities for each orientation and location.

Simple cells come with even-symmetric and odd-symmetric receptive fields. A new functional role for even-symmetric and odd-symmetric simple cells was identified through the analysis of a problem with edge localization that results from convolving ON and OFF

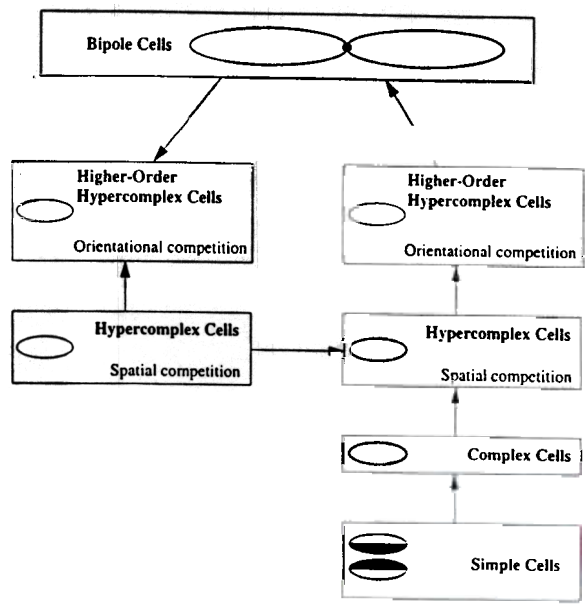

FIGURE 4. Schematic of a single-scale BCS model, adapted from Grossberg et al. (1995). The simple, complex cell, and hypercomplex stages are often referred to as the Static Oriented Contrast filter (SOC filter), while the subsequent levels, including their feedback pathways, are commonly referred to as the Static Oriented Competitive Cooperative loop (SOCC loop).

LGN output signals with simple cell receptive fields. For example, at a dark/light vertical edge, dark/light vertically oriented simple cells have their peak in the center of the edge. However, light/dark vertical oriented cells respond to this input pattern with two smaller peaks, one on either side of the luminance border. An example of this phenomenon is shown in Figure 5, along with the input pattern, the LGN ON and OFF responses, and the simple oriented detectors responses. This input pattern represents a one-dimensional horizontal slice through the right eye view presented in Figure 3. The final frame of Figure 5 displays the results of summing the half-wave rectified outputs of oppositely polarized simple cells. Three peaks are registered at each edge. Without further processing, these extra peaks could cause problems in the binocular disparity-matching process.

This problem is solved by pooling responses from odd-symmetric and even-symmetric simple cells at the complex cell level. As illustrated in Figure 6, pooling responses of both odd and even symmetric simple cells of both contrast polarities solves the triple peak problem for both light/dark and dark/light boundaries, thereby achieving good boundary localization while removing spurious peaks in the disparity-matching process. Pollen and his colleagues have previously reported on the coexistence of cells with even and odd symmetric receptive fields in the cat visual cortex (Pollen et al., 1985; Pollen & Ronner, 1981, 1982). The present analysis suggests a new functional reason for the coexistence of even-symmetric and odd-symmetric simple cells.

Even and odd symmetric simple cell receptive fields

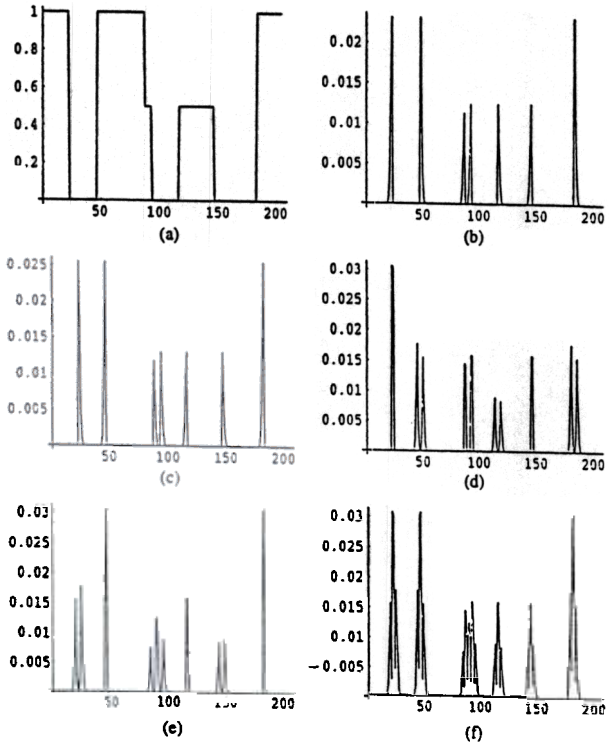

FIGURE 5. One-dimensional slices through the activity caused by the right eye view of the da Vinci stereogram for: (a) the input; (b) the ON channel; (c) the OFF channel; (d) the odd symmetric light/dark oriented simple cells of size 1; (e) the odd symmetric dark/light oriented simple cells of size 1; (f) the result of summing the dark/light and light/dark simple cell responses. Note the triple peaks evident in (f). (Within each plot the horizontal axis represents space while the vertical axis measures activity strength).

centered at location ij of orientation k were defined using the even and odd Gabor kernels:

$$S_{ijk}^{\text{odd}+/-} = \left[\sum_{(p,q)} s_{pqk}^{\text{odd}} X_{i-p,j-q}^{+/-} - \sum_{(p,q)} s_{pqk}^{\text{odd}} X_{i+p,j+q}^{-/+} \right]^{+} \tag{12}$$

and

$$S_{ijk}^{\text{even}+/-} = \left[\sum_{(p,q)} s_{pqk}^{\text{even}} X_{i-p,j-q}^{+/-} - \sum_{(p,q)} s_{pqk}^{\text{even}} X_{i+p,j+q}^{-/+} \right] \tag{13}$$

$$s_{pqk}^{\text{odd}} = A \sin\left(\frac{2\Pi k}{T}\right) \exp\left[-\frac{1}{2}\left(\frac{p^2}{\sigma_{pk}^2} + \frac{q^2}{\sigma_{qk}^2} \right) \right] \tag{14}$$

and

$$s_{pqk}^{\text{even}} = A \cos\left(\frac{2\Pi k}{T}\right) \exp\left[-\frac{1}{2}\left(\frac{p^2}{\sigma_{pk}^2} + \frac{q^2}{\sigma_{qk}^2} \right) \right] \tag{15}$$

with $[x]^+ = \max(0, x)$, $-S_1 \le p,\ q \le S_1$ (Scale 1) and $-S_2 \le p,\ q \le S_2$ (Scale 2).

In eqns (12) and (13), $S_{ijk}^{\text{odd}+}$ is the positive-polarity odd-symmetric simple cell response at location ij and orientation k; $S_{ijk}^{\text{even}+}$ is the positive-polarity even-symmetric simple cell response; and $S_{ijk}^{\text{odd}-}$ and $S_{ijk}^{\text{even}-}$ define the negative-polarity simple cell responses. In eqns (14) and (15), A is a scale constant (1); T determines the periodicity of the simple cell (scale 1: even $= \pi$, odd $= 2\pi$; scale 2: even $= \pi$, odd $= 3\pi$); S_1 and S_2 define the size of the simple cell input field (4, 6); σ_{pk} and σ_{qk} define the standard deviations of the x and y dimensions of simple cell input field for orientation k [scale 1: vertical ($\sigma_p = 2$, $\sigma_q = 1.5$); horizontal ($\sigma_p = 1.5$, $\sigma_q = 2$); scale 2: vertical ($\sigma_p = 2.5$, $\sigma_q = 2$); horizontal ($\sigma_p = 2$, $\sigma_q = 2.5$)].

4.3. Complex Cells

The complex cell level contains the first binocularly activated cells of the model. For definiteness, we assume that the da Vinci display is viewed in such a way that only crossed, zero disparity, and monocular cell pools are activated. To understand how binocular matching occurs in the presence of half-occluded regions, it was noted by Grossberg (1994) that the binocular filter needs to realize a type of size-disparity correlation (Richards & Kaye, 1974; Tyler, 1975, 1983; Kulikowski, 1978; Schor & Tyler, 1981; Schor & Wood, 1983; Schor et al., 1984) whereby cells with larger receptive field sizes, or spatial scales, can binocularly fuse a larger range of binocular disparities than can cells with smaller receptive field sizes. Hence, larger cells can fuse a range of both small and large disparities. This hypothesis is in contrast to alternative approaches (e.g., Marr & Poggio, 1979) wherein large spatial scale cells code only large disparities and each spatial scale can fuse only a limited range of disparities. Julesz and Schumer (1978) reviewed psychophysical data that are inconsistent with the Marr–Poggio model, but which support the FACADE theory conception. This scheme can explain how binocular percepts are binocularly fused at a larger spatial scale, but rivalrous at a smaller spatial scale (Julesz & Miller, 1975; Kulikowski, 1978) since the large spatial scale can fuse a larger disparity range than the smaller spatial scale.

The binocular filter contains a number of new computational features. Each spatial scale is composed of distinct cell pools, with each pool coding a different disparity, or connected range of disparities. Although each simple cell is sensitive to just one contrast polarity, complex cells typically pool signals from opposite contrast polarities. Sandwiched between these properties, the binocular matching process that occurs within the simple-to-complex cell filter facilitates binocular matching of contours with the same contrast polarity. These properties are achieved using the filter described in Figure 7. Here, outputs from like-polarity simple cells first undergo excitatory matching, whereas opposite polarity matches inhibit each other. Net signals are half-wave

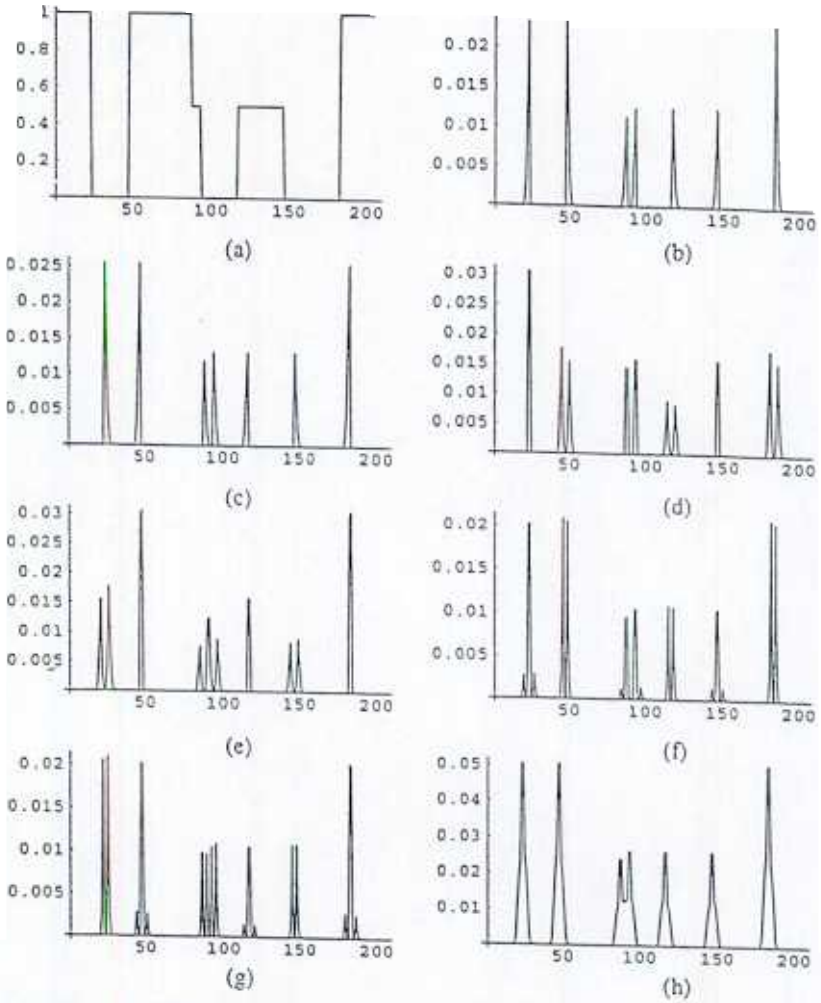

FIGURE 6. One-dimensional slices through the activity caused by the right eye view of the da Vinci stereogram for: (a) the input; (b) the ON channel; (c) the OFF channel; (d) the odd symmetric light/dark oriented simple cells of size 1; (e) the odd symmetric dark/light oriented simple cells of size 1; (f) the even symmetric light/dark oriented simple cells of size 1; (g) the even symmetric dark/light oriented simple cells of size 1; (h) the result of summing the rectified output of the odd and even simple cells. Note that unimodal peaks are generated by each edge. (Once again the horizontal axis represents space while the vertical axis measures activity strength.)

rectified and pooled across polarities at a target complex cell. Each complex cell receives this sort of input from both even and odd symmetric simple cell receptive fields. As a result, like-contrast polarity boundaries may be binocularly fused while opposite-contrast polarity boundaries are mutually rivalrous. This circuit helps to explain the difficulty that observers experience in fusing correlated random dot patterns of opposite polarity (von Helmholtz, 1910/1925; Anderson & Nakayama, 1994; Harris & Parker, 1995) without contradicting the fact that complex cells respond to both polarities.

Adding half-wave rectified outputs of oppositely polarized simple cells at complex cells has the net effect that the complex cells full-wave rectify the input image. This hypothesis has become commonplace in models of texture segregation (Grossberg & Mingolla, 1985b; Chubb & Sperling, 1989; Sutter et al., 1989). Here it emerges as a consequence of binocular matching.

The disparity tuning of the complex cells is obtained in two ways. First, each cell receives activity from two horizontally shifted left and right eye simple cells. These inputs are weighted so that each cell fires maximally to a boundary presented at a given disparity. It is assumed that the monocular weights which define the binocular receptive field of the complex cell, are pruned during development so that different complex cells code distinct disparities. Grunewald and Grossberg (1997) have simulated how this process of disparity tuning may self-organize during development to create the binocular cell interactions that were introduced in the present research. In the present simulations, these kernels are one-dimensional, since physiological results suggest that disparity-tuned cells are sensitive to horizontal disparity and are suppressed by even small amounts of vertical disparity (e.g., one arc-minute or so: Poggio, 1984; but see also Gonzalez et al., 1993). When a complex cell receives its ideal pattern of cortical activity from the simple cell level, it registers a high pattern match and is strongly activated. As the cortical pattern is changed from its ideal input (for example, by introducing more or

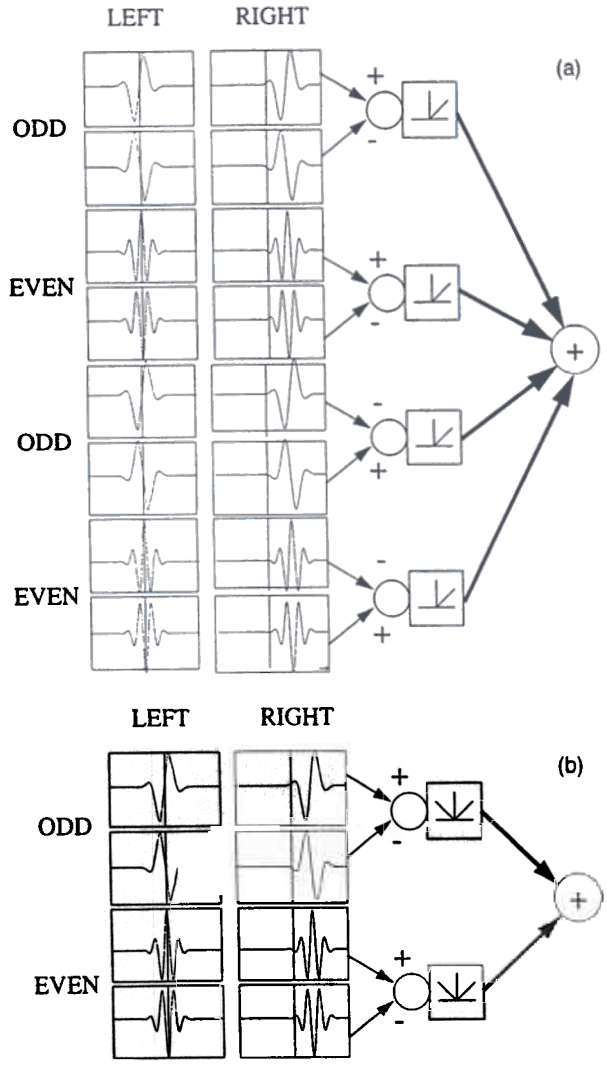

FIGURE 7. Binocular simple-to-complex cell disparity filter: (a) each cell matches a like polarity pair of odd and even receptive fields. After matching, opposite polarity matched pairs inhibit each other and the result is half-wave rectified. This ensures that the complex cell will fire strongly to matches of either polarity. Finally the matched even and odd donations are summed. (b) An equivalent circuit with full-wave rectification.

less horizontal disparity), the cell registers a less perfect match and is not as strongly activated.

Inhibitory interactions between the complex cells represent the second mechanism by which they achieve disparity tuning (Figure 8). These inhibitory interactions suppress false binocular matches as well as the monocular representations that occur within binocularly fused regions. The developmental principle that *cells that fire together wire together* is implemented to help explain how the wiring pattern of these inhibitory connections develops. In particular, only cells that receive input from a common simple cell compete. There is a nice temporal correlation between the development of ocular dominance columns, the pruning of horizontal excitatory and inhibitory connections, and the onset of stereopsis in the monkey visual cortex (Shimojo et al., 1986; Schatz, 1992; Katz & Callaway, 1992; Löwel & Singer, 1992). We postulate that this pruning of connections results in selective inhibitory connections between cells which share a common simple cell input. In other words, cells that responded to a common input prior to the onset of stereopsis stay wired together in an inhibitory manner.

Additional evidence for such suppressive interactions come from the psychophysical studies of McKee and colleagues (McKee et al., 1990; McKee & Harrad, 1993) which found that monocular positional information is corrupted if a disparate stimulus is presented to the other eye. They suggest "a kind of local competition among neural units serving the same retinal location. The neural units with a strong response...inhibit the neurons with a weak response" (McKee & Harrad, 1993, p. 1646). These inhibitory interactions are generated only when a cell is sufficiently active (over its firing threshold). This property correlates with the idea of there being a fusion threshold, which has recently gained considerable support from the work of McKee and colleagues (McKee, 1993; McKee et al., 1994) on dichoptic masking, and from the work of Liu et al. (1992) on stereopsis at low contrast. In both cases, binocular fusion was found to be contrast-sensitive. The binocular matching process of the current model is also contrast-sensitive. A complex cell must exceed its firing threshold if it is to be registered by the subsequent processing stages. When a complex cell exceeds its firing threshold, it engages the inhibitory interactions described above to kill off all weaker matches.

The competition between cells coding different disparities at each position enables FACADE theory to circumvent the uniqueness principle of classical stereopsis algorithms. The uniqueness principle, simply stated, is the idea that each point in the left view of a stereogram can have one and only one match in the right view. A recent study by McKee et al. (1994) suggests that in certain circumstances the uniqueness principle is indeed violated. These data, among others, have been simulated using the present model (McLoughlin & Grossberg, 1995, 1997).

In the current simulation, the smaller spatial scale contains four separate disparity-sensitive pools of cells. One pool corresponds to zero disparity (disp00), another pool to the disparity of the far wall and window (disp01), another to an intermediate disparity (disp02), and the final cell pool to the disparity of the near surface (disp03). The larger spatial scale contains five pools of disparity tuned cells. The first four groups are tuned to the same disparities as the smaller scale (disp10, disp11, disp12, disp13). The final pool is tuned to a larger disparity than that of the near wall (disp14).

Poggio et al. (1988) have proposed that disparity selectivity is pooled into six groups of cells. Tuned excitatory (TO) and tuned inhibitory (TI) cells, which respond sharply to stimuli at zero disparity (in excitatory and

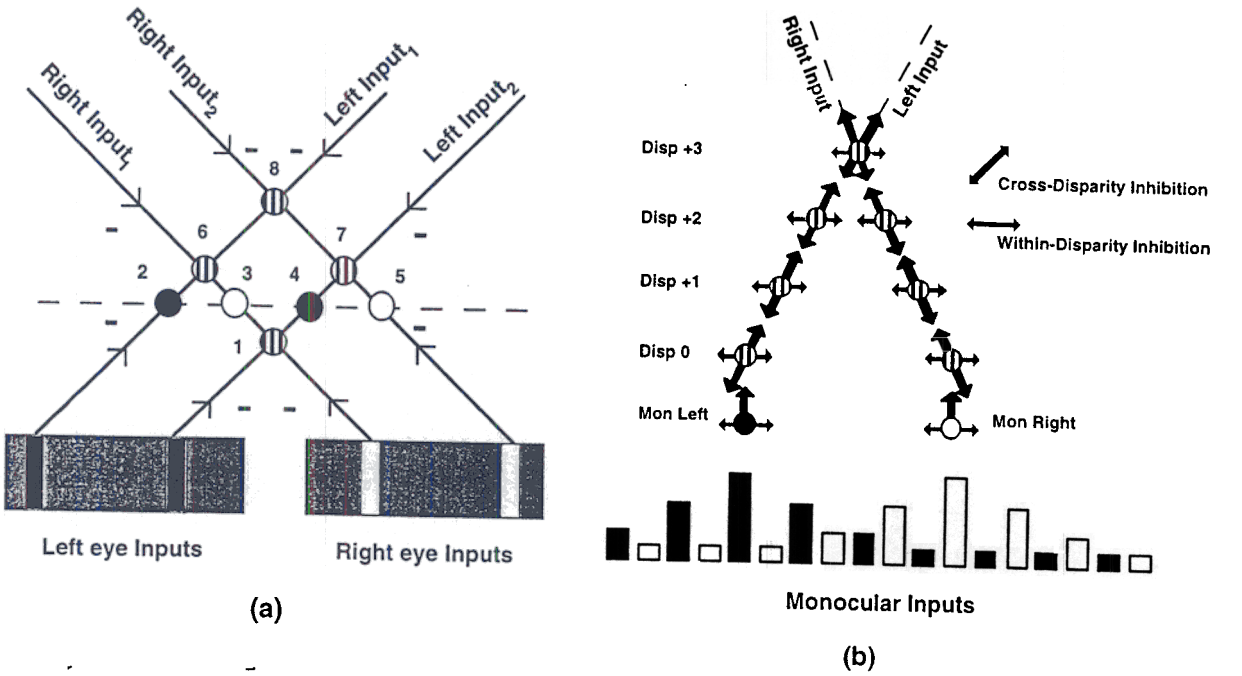

FIGURE 8. Inhibition occurs only between cells that attempt to code the same input. This principle has been called "inhibition along line-of-sight". (a) Keplerian schematic of this matching process. Left and right eye inputs are represented as black and white bars and are projected along hypothetical input lines representing their various binocular and monocular combinations. Zero disparity is delineated by the broken horizontal line. Monocular left eye complex cells are displayed as filled (black) circles, monocular right eye complex cells are displayed as unfilled (white) circles, and binocular complex cells are displayed as striped (black and white) circles. Cells at locations 2–5 attempt to encode the incoming signals as monocular events. (Zero disparity binocular complex cells also present at these locations are omitted for clarity.) Cells at location 1 and locations 6–8, attempt to encode the incoming signals at various binocular disparities. Inhibition occurs between all cells residing on the common input lines: left $input_1$; left $input_2$; right $input_1$; and right $input_2$. This figure oversimplifies the problem of binocular matching, however, as it neglects cells other than those at locations 1–8 which also attempt to encode the inputs. (b) A more complete description of the selection process. Cells are active at each possible binocular disparity (including zero disparity) along the input signal lines. These additional cells, not present in (a), receive strong input from one eye, with little or no input from the other, and need to be suppressed. *Cross-disparity* inhibition occurs between all cells that receive input from the same input line and helps to localize the complex cells response in depth. This includes the monocular complex cells (the completely filled and unfilled circles). *Within-disparity* inhibition occurs within each complex cell pool and helps to localize the complex cells response to inputs within a given disparity plane.

inhibitory fashions), near (NE) and far (FA) cells, which respond to crossed and uncrossed disparities of relatively large magnitude in an excitatory manner, and tuned near (TN) and tuned far (TF) cells, which respond to smaller crossed and uncrossed disparities in an excitatory manner. Both TO and TI cells have sharper tuning functions than the other cell pools, NE and FA cells having the most shallow falloff. Richards (1971) had previously predicted the existence of such pools based on clinical studies of humans.

The disparity-tuned filter of the model is consistent with Poggio's physiological results on pooling. The distribution of disparity cells within the filter is biased towards zero disparity. In other words, the disparity difference between each pool increases from disp00 to disp01, and so on. This increase in distance between pools translates into a shallower falloff for pools tuned to larger disparities. This filter does, however, make use of more pools than Poggio lists. More cell pools are required to ensure that the fused surfaces are represented correctly. Physiological evidence from the cat (LeVay & Voight, 1988) and psychophysical evidence from adaptation studies in humans (Stevenson et al., 1992) suggest that there may, in fact, be a continuum of pools in these species.

As in Grossberg (1994), cells sensitive to horizontal boundaries are projected in parallel to all disparity-sensitive pools of cells in order to help complete connected boundaries. Strictly monocular cells were not incorporated into the Grossberg (1994) model, which distinguished nonzero disparity cells (for computing horizontal disparities) from near-zero disparity cells (for registering horizontal and monocularly viewed features). The present extension of the model posits monocular cells that are distinct from near-zero disparity cells. Separate pools of left and right eye monocular units represent unfused monocular stimuli. Unfused monocular stimuli may be grouped with fused binocular stimuli at any depth plane, hence the need for distinct monocular cell pools. These monocular cell pool activities (monL0, monR0) and (monL1, monR1) are topographically added to each of the binocular disparity-sensitive cell pools

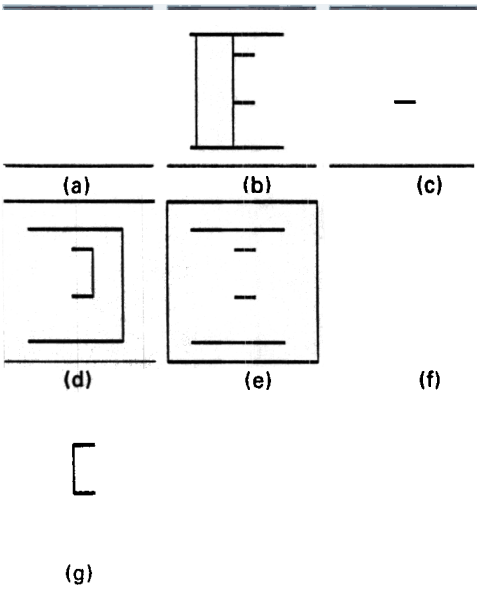

FIGURE 9. Output of the disparity filter for the large spatial scale: (a) No vertical boundaries are represented in the disp14 cell pool. (b) The near surface boundary representation (AB) resides in the disp13 cell pool. (c) No vertical boundaries are fused within the disp12 cell pool. (d) The right-hand side of the far wall and window are represented in disp11. (e) The picture frame is correctly fused at zero disparity in the disp10 cell pool. (f) No vertical monocular boundaries remain unfused in the left eye monocular cell pool. (g) Finally, the left-hand side of the far window passes through the disparity filter without being suppressed. It is represented in the monocular right eye cell pool.

(disp00–disp14) in a scale-specific manner. This convergence of signal pathways occurs between the disparity filter and model hypercomplex cells. McKee and Harrad (1993) have shown that subjects are unable to access monocular information when it is part of a fused binocular percept. They named this phenomenon "fusional suppression", since the fusional process appears to suppress the monocular representations, as in the current model.

Figure 9 summarizes the output of the large spatial scale disparity filter after processing the da Vinci stereogram. The vertically and horizontally tuned cell activity profiles are displayed together to aid in comprehension. The large spatial scale does an excellent job of segmenting the scene into its relative depth planes. The cell pool tuned to the largest disparity (disp14) remains inactivated [Figure 9(a)]. No vertical boundaries are fused at this depth. Disp13 cells fuse the vertical boundaries of the near lefthand surface [Figure 9(b)]. In so doing, they suppress all false matches associated with these boundaries within the large spatial scale via mutual complex cell inhibition. Hence, no vertical boundaries corresponding to the near surface are registered in Figure 9(a), or in Figure 9(c)–(g). The next disparity-sensitive vertically oriented cell pool, disp12, is also inactive [Figure 9(c)]. Once again no surfaces exist within this depth plane of the stimulus. Disp11 cells fuse the vertical boundaries of the far wall and window [Figure 9(d)]. The left-hand side of the window is presented monocularly due to occlusion and so is registered in the monocular right (MonR1) cell pool [Figure 9(g)]. Finally the vertical boundaries of the background are fused at zero disparity (disp10) in Figure 9(e), and the monocular left (MonL1) vertical cell pool is silent [Figure 9(f)].

Figure 10 presents the outputs of the small spatial scale's disparity filter obtained after processing the da Vinci stereogram. Only four disparity selective pools exist within this spatial scale. Due to the size-disparity correlation, the small spatial scale is not capable of fusing surfaces present at disparities greater than disp03. This restriction introduces the possibility of a surface being fused within a large spatial scale's disparity filter while remaining outside the fusional range of smaller spatial scales. If this phenomenon were to occur, one would expect either a fuzziness at the boundaries of the large disparity surface or a combination of rivalry and fusion at the same spatial location (Grossberg, 1987b, 1994). Both phenomena have been reported for suitable stimuli (McKee, personal communication; Kulikowski, 1978).

However, in the case of the da Vinci stereogram, the nearest surface is within the fusional range of the small spatial scale and is represented at the same disparity as in the large spatial scale [disp03, Figure 10(a)]. As in the large spatial scale, no vertical boundaries are registered in the monocular left [MonL0, Figure 10(e)] or disp02 cell pools [Figure 10(b)]. Vertical boundaries of the far wall and window are represented at disp01 [Figure 10(c)], the background is registered at disp00 [Figure 10(d)], and the right-hand side of the window is represented in the monocular right cell pool [Figure 10(f)].

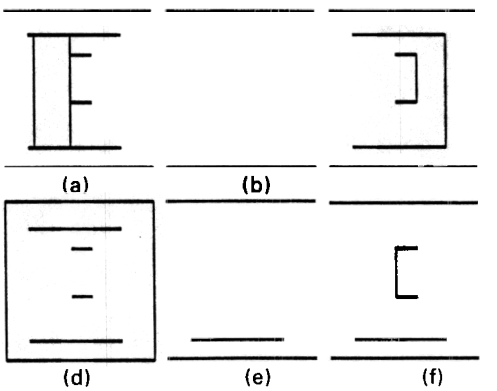

FIGURE 10. Output of the disparity filter for the small spatial scale. (a) As in the larger scale, the disparity sensitive disp03 cell pool fuses the near surface (AB). (b) No vertical boundaries are fused at disp02. (c) The disp01 representation once again codes the right-hand side of the far wall and window. (d) The picture frame is fused at zero disparity, namely disp00. (e) No vertical boundaries from the left eye are unfused. (f) Additionally, as in the larger spatial scale, the left-hand side of the far window is unfused and represented in the monocular right eye cell pool.

The dynamics of the complex cell stage are defined mathematically as follows:

$$\frac{dc_{ijkd}}{dt} = -\alpha_2 c_{ijkd} + (U_2 - c_{ijkd})M_{ijkd} - (c_{ijkd} + L_2)\sum_{(p,e)} h(c_{i(j+p)ke})D_{i(j+p)kde} \quad (16)$$

In eqn (16), α_2 is the decay rate (0.01); U_2 bounds a cell's maximum activity (1); M_{ijkd} defines the simple-to-complex cell filter (defined below); L_2 bounds the cell's minimum activity (1); $\Sigma h(c_{i(j+p)e})$ is the total inhibitory signal from cells that receive the same simple cell (s_{ijk}) inputs tuned to disparity e. The signal function h is a half-wave rectified analog signal with threshold Γ_e:

$$h(c_{ijke}) = \begin{cases} c_{ijke} - \Gamma_e & c_{ijke} > \Gamma_e \\ 0 & \text{otherwise} \end{cases} \quad (17)$$

Term $D_{i(j+p)kde}$ in eqn (16) represents the inhibitory connections between the cells that code disparity e and the cells that code disparity d. These are assumed to be pruned by adaptive inhibitory learning during development so that only cells that receive a common input are connected. In particular, $D_{i(j+p)kde} \neq D_{i(j+p)ked}$.

The simple-to-complex cell filter combines simple cell output signals in the following way:

$$M_{ijkd} = \beta_2 \left| \left(\sum_{l=1}^{L} W_{dlk}^{\text{even}} S_{i,j+l,k}^{\text{even}+} + \sum_{r=1}^{R} W_{drk}^{\text{even}} S_{i,j+r,k}^{\text{even}+} \right) - \left(\sum_{l=1}^{L} W_{dlk}^{\text{even}} S_{i,j+l,k}^{\text{even}-} + \sum_{r=1}^{R} W_{drk}^{\text{even}} S_{i,j+r,k}^{\text{even}-} \right) \right| + \left| \left(\sum_{l=1}^{L} W_{dlk}^{\text{odd}} S_{i,j+l,k}^{\text{odd}+} + \sum_{r=1}^{R} W_{drk}^{\text{odd}} S_{i,j+r,k}^{\text{odd}+} \right) - \left(\sum_{l=1}^{L} W_{dlk}^{\text{odd}} S_{i,j+l,k}^{\text{odd}-} + \sum_{r=1}^{R} W_{drk}^{\text{odd}} S_{i,j+r,k}^{\text{odd}-} \right) \right| \quad (18)$$

In eqn (18), M_{ijkd} is the total input to the complex cell centered on location ij, of orientation k, and tuned to disparity d; $W_{d(l/r)k}^{\text{even/odd}}$ are weighting functions corresponding to disparity d, distance (l/r), and orientation k; $S_{ijk}^{\text{even}+/-}$ is the even symmetric simple cell input of positive or negative polarity at location ij and orientation k; $S_{ijk}^{\text{odd}+/-}$ is the corresponding odd-symmetric simple cell input; L, R correspond to the size of the simple cell input fields in the left and right eye (scale 1, L: 20, R: 20; scale 2, L: 28, R: 28). By eqn (18), complex cells that receive input from disparity d pools of opposite-polarity simple cells will not be activated, while those that receive input from disparity d pools of like-polarity simple cells will be strongly activated. $\beta_2 = 15$.

In the current implementation, the disparate monocular weights $W_{d(l/r)k}^{\text{odd}}$ and $W_{d(l/r)k}^{\text{even}}$ in eqn (18) are set proportional to the strength of the simple cell responses to a binocular step edge presented at disparity d. Thus

$$W_{drk}^{\text{odd/even}} = \frac{S_{i,j+r,k}^{\text{odd/even}}}{\sum_{r=1}^{R} S_{i,j+r,k}^{\text{odd/even}}}, \quad W_{dlk}^{\text{odd/even}} = \frac{S_{i,j+l,k}^{\text{odd/even}}}{\sum_{l=1}^{L} S_{i,j+l,k}^{\text{odd/even}}} \quad (19)$$

where as in eqns (8)–(13)

$$S_{ijk}^{\text{odd/even}+} = \left[\sum_{(p,q)} s_{pqk}^{\text{odd/even}} X_{i-p,j-q}^{+} - \sum_{(p,q)} s_{pqk}^{\text{odd/even}} X_{i+p,j+q}^{-} \right]^{+},$$

$$S_{ijk}^{\text{odd/even}-} = \left[\sum_{(p,q)} s_{pqk}^{\text{odd/even}} X_{i-p,j-q}^{-} - \sum_{(p,q)} s_{pqk}^{\text{odd/even}} X_{i+p,j+q}^{+} \right]^{+}$$

and

$$X_{ij}^{+} = \frac{\left[(U_1 + L_1)\left(\sum_{(p,q)} C_{pq} - S_{pq} \right) I_{i+p,j+q} \right]^{+}}{\alpha_1 + \sum_{(p,q)} (C_{pq} + S_{pq}) I_{i+p,j+q}}$$

with

$$I_{ij} = \begin{cases} 0.0 & \text{if } j < \frac{N}{2} \\ 1.0 & \text{if } j \geq \frac{N}{2} \end{cases}$$

Terms X_{ij}^{-} in eqn (20) are defined similarly. Finally, the inhibitory connections $D_{i(j+p)kde}$ between complex cells of the same spatial scale that are tuned to different disparities were implemented as follows:

$$D_{i(j+p)kde} = A_{de} \exp(-\mu_{de}(p + s_{de})^2) \text{ for } d \neq e \quad (23)$$

and

$$D_{i(j+p)kdd} = A_{dd}[\exp(-\mu_{dd}^{c} p^2) - \exp(-\mu_{dd}^{s} p^2)] \quad (24)$$

In eqns (23) and (24), A_{de} corresponds to a scale parameter for the strength of inhibition between disparity d and disparity e (see Table 1–2) μ_{de} is the spatial parameter that determines the width of the inhibitory gaussian between cells coding disparities d and e (see

TABLE 1

All tables demonstrate new binocular interactions introduced in Grossberg (1994). Scale values A_{de} for scale 1

	A_{mon}	A_0	A_1	A_2	A_3
A_{mon}					
A_0					
A_1					
A_2					
A_3					

TABLE 2
Scale values A_{de} for scale 2

	A_{mon}	A_0	A_1	A_2	A_3	A_4
A_{mon}						
A_0						
A_1						
A_2						
A_3						
A_4						

TABLE 3
Shift parameters s_{de} for scale 1

	s_{monL}	s_{monR}	s_0	s_1	s_2	s_3
s_{monL}		—				
s_{monR}		0				
s_0		0				
s_1		+1				
s_2		+2				
s_3		+5				

TABLE 4
Shift parameters s_{de} for scale 2

	s_{monL}	s_{monR}	s_0	s_i	s_2	s_3	s_4
s_{monL}		—					
s_{monR}		0					
s_0		0					
s_1		+1					
s_2		+2					
s_3		+5					
s_4		+7					

below); s_{de} determines the shift in the center of the Gaussian between disparity d and disparity e (see Tables 3–4). A_{dd} is the scale parameter for the spatial inhibition kernel D_{ipkdd}; μ_d^c determines the width of the excitatory center for the within disparity interaction; and μ_d^s determines the width of the inhibitory surround for the within disparity interaction. For scale 1: $\mu_{dmon} = 0.075$, $\mu_{de} = 0.05$, $\mu_{mone} = 0.075$; $(d \in \{mon, 0, 1\})$ $\mu_d^c = 0.015$, $\mu_d^s = 0.5$; $(d \in \{2, 3\})$ $\mu_d^c = 0.005$, $\mu_d^s = 0.5$. For scale 2:scale 2: $\mu_{dmon} = 0.075$, $\mu_{de} = 0.05$, $\mu_{mone} = 0.075$; $(d \in \{mon, 0, 1\})$ $\mu_d^c = 0.015$, $\mu_d^s = 0.5$; $(d \in \{2, 3, 4\})$ $\mu_d^c = 0.005$, $\mu_d^s = 0.5$.

Cells sensitive to horizontal boundaries are spatially sharpened using the following on-center off-surround network:

$$\frac{dc_{ijk}}{dt} = -\alpha_3 c_{ijk} + (U_3 - c_{ijk}) \sum_p C_{i(j+p)} M_{i(j+p)k} - (L_3 + c_{ijk}) \sum_p E_{i(j+p)} M_{i(j+p)k} \quad (25)$$

The Gaussian center C_{ij} and surround E_{ij} kernels are defined as follows:

$$C_{i(j+p)} = A_1 \exp[-\mu^c p^2], \text{ and } E_{i(j+p)} = A_2 \exp[-\mu^s p^2] \quad (26)$$

In eqn (25), c_{ijk} is the activity of the horizontal complex cell at position ij (all disparities are the same); α_3 is the decay parameter (0.1); M_{ijk} is defined in eqn (18); p determines the size of this spatial interaction (15); C_{ij} and E_{ij} are the center and surround Gaussian kernels ($\mu_c = 6.9$, $\mu_s = 0.69$, $A_1 = 2$, $A_2 = 1$); U_3 and L_3 bound the activity of the horizontally tuned cells (1, 1). The half-wave rectified, steady-state output of eqn (25) is

$$C_{ijkd} = \frac{\sum_p [(U_3 C_{i(j+p)} - L_3 E_{i(j+p)}) M_{i(j+p)k}]^+}{\alpha_3 + \sum_p (C_{i(j+p)} + E_{i(j+p)}) M_{i(j+p)k}} \quad (27)$$

4.4. Hypercomplex Cells: Spatial Competition

The rectified output from each complex cell pool excites a like-oriented hypercomplex cell corresponding to its location and inhibits like-oriented hypercomplex cells corresponding to nearby locations. This spatial inhibition occurs in a center-surround manner between cells coding the same scale, disparity and orientation. It models the end stopping process, whereby hypercomplex cell receptive fields are derived from interacting complex cell output signals (Hubel and Wiesel, 1965; Orban et al., 1979).

Each hypercomplex cell activity y_{ijkd} also receives feedback from the final stage of the cooperative grouping network (called the SOCC Loop) and has a baseline or tonic activity level T. Hypercomplex cell dynamics are defined by

$$\frac{dy_{ijkd}}{dt} = -\alpha_4 y_{ikd} + (U_4 - y_{ikd}) C_4 - (y_{ijkd} + L_4) E_4 + T + g(O_{ijkd}) \quad (28)$$

The center C_4 and surround signals E_4 are defined as follows:

$$C_4 = \sum_{(h,v)} \Phi_{hv} C_{i+h,j+v,k,d}$$

with

$$\Phi_{hv} = \frac{A \exp\left[-\frac{1}{2}\left(\frac{h^2+v^2}{\sigma_c^2}\right)\right]}{\sum_{(h,v)} \exp\left[-\frac{1}{2}\left(\frac{h^2+v^2}{\sigma_c^2}\right)\right]} \quad (29)$$

and

$$E_4 = \sum_{(h,v)} E_{hv} C_{i+h,j+v,k,d}$$

with

$$E_{hv} = \frac{B \exp\left[-\frac{1}{2}\left(\frac{h^2+v^2}{\sigma_s^2}\right)\right]}{\sum_{(h,v)} \exp\left[-\frac{1}{2}\left(\frac{h^2+v^2}{\sigma_s^2}\right)\right]} \quad (30)$$

where $-P_1^h \le h,\ v \le P_1^h$ (scale 1) and $-P_2^h \le h, v \le P_2^h$ (scale 2).

In eqn (28), α_4 is the decay parameter (0.1); U_4 and $-L_4$ are the upper and lower bounds of hypercomplex cell activity (1, −1); T is the tonic activity of the cell (0.0000189); $g(O_{ijkd})$ is the feedback activity from the SOCC loop, where $g(x) = Gx$, with G a scale constant (1). In eqns (29) and (30), C_{ijkd} is the complex cell activity at location ij, orientation k and disparity d; P_1^h and P_2^h define the size of the spatial interaction for scale 1 and scale 2, respectively (4, 4); A and B are scale constants (1, 1); and σ_c and σ_s are the standard deviations of the center and surround Gaussians (scale 1: 0.15, 1; scale 2: 0.15, 1). Solving eqn (28) at equilibrium and half-wave rectifying the result yields the steady-state hypercomplex cell output signal:

$$Y_{ijkd} = \frac{\left[\sum_{(h,v)} (U_4\Phi_{hv} - L_4E_{hv})C_{i+h,j+v,k,d} + T + g(O_{ijkd})\right]^+}{\alpha_4 + \sum_{(h,v)} (\Phi_{hv} + E_{hv})C_{i+h,j+v,k,d}} \tag{31}$$

4.5. Hypercomplex Cells: Orientational Competition

The next stage of hypercomplex cells compete across orientation at each position, with the greatest inhibition occurring between cells that code perpendicular orientations. This competition is a push–pull opponent process whereby, if one orientation is excited (or inhibited), then the cell coding the perpendicular orientation at the same spatial location is inhibited (or excited by disinhibition). Spatial competition causes vertically tuned cells just beyond the end of a thin vertical line to be inhibited. Horizontally tuned hypercomplex cells that code the same spatial position are then disinhibited by orientational competition. This mechanism of completing line ends, called endcutting, helps to explain, for example, how illusory contours can be formed perpendicular, or oblique, to line ends (Grossberg & Mingolla, 1985b). Such endcuts may be grouped together by subsequent stages of the SOCC loop to form an illusory contour, as in the Ehrenstein illusion; see Gove et al. (1995) for a simulation. Cross-orientation inhibition occurs between vertically and horizontally tuned cells in the present simulations. These cells also receive feedback from the monocular FIDOs (Figure 2, pathways 7).

These properties of hypercomplex cell activities n_{ijkd} were modeled as follows:

$$\frac{dn_{ijkd}}{dt} = -\alpha_5 n_{ijkd} + (U_5 - n_{ijkd})C_5 - (n_{ijkd} + L_5)E_5 - \sum_{e \ge d} F(v_{ijke}) \tag{32}$$

$$C_5 = \sum_r C_{kr}Y_{ijrd} \text{ with } C_{kr} = \frac{c}{2\pi\sigma_c^2}\exp\left[-\frac{1}{2}\left(\frac{r-k}{\sigma_c^2}\right)^2\right] \tag{33}$$

and

$$E_5 = \sum_r E_{kr}Y_{ijrd} \text{ with } E_{kr} = \frac{s}{2\pi\sigma_s^2}\exp\left[-\frac{1}{2}\left(\frac{r-k}{\sigma_s^2}\right)^2\right] \tag{34}$$

In eqn (32), α_5 is the decay parameter (1); U_5 and $-L_5$ are the upper and lower bounds of hypercomplex cell activity (1, −1); $F(v_{ijke})$ corresponds to the FCS to BCS inhibitory feedback pathways where $F(x) = Fx$, with F a scale constant (10). In eqns (33) and (34), Y_{ijkd} is the hypercomplex cell output (31) from the first competitive stage at location ij, orientation k and disparity d; c and s are scale constants of the center and surround Gaussians (1, 1.5); σ_c and σ_s are the standard deviations of the center and surround Gaussians (scale 1: 0.5, 0.75; scale 2: 0.5, 0.75); and r and k denote orientation indices ($r \in \{0, 1\}, k \in \{0, 1\}$). Solving at equilibrium and half-wave rectifying the result yields the steady-state higher-order hypercomplex cell output signal:

$$N_{ijkd} = \frac{\left[\sum_r (U_5C_{kr} - L_5E_{kr})Y_{ijkd} - \sum_{e \ge d} F(v_{ijke})\right]^+}{\alpha_5 + \sum (C_{kr} + E_{kr})Y_{ijkd}} \tag{35}$$

4.6. Cooperative Bipole Cells

Outputs from the second hypercomplex cell stage are fed into a long-range cooperative boundary completion process which is responsible for the long-range grouping capabilities of the BCS (Figure 4). The long-range cooperation cells that receive this input are called bipole cells, so called because their receptive field is composed of two oriented lobes. Each lobe receives input from a range of orientations and positions along the oriented axis of the lobe. If both bipole lobes (or the cell itself) receive large enough inputs, the cell will fire. This restriction ensures that bipole cells do not extend boundaries beyond line ends unless there is evidence for such a linkage, for example from a second aligned line end. Bipole cells hereby behave like statistical AND gates. Bipole cell outputs are fed back to the hypercomplex cells to initiate boundary completion. The existence of bipole cells was predicted (Grossberg, 1984; Cohen & Grossberg, 1984) at around the time von der Heydt et al. (1984) reported analogous cell properties in monkey visual area V2.

Bipole cells receive inputs from all cells coding the same relative depth from the observer across all hypercomplex cell spatial scales. These converging signals

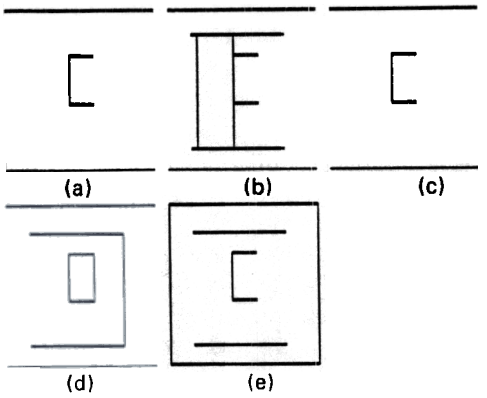

FIGURE 11. Open loop response of the bipole cells. (a) The disp04 representation contains only the monocularly viewed left-hand side of the far window. No completely enclosed surfaces exist at this depth plane. (b) The near surface (AB) is represented by a completely enclosed boundary structure. This completely enclosed structure traps filling-in signals within the monocular FCS filling-in domains associated with this depth. (c) Once again no completely enclosed surfaces exist at disp02. (d) The window remains the only enclosed surface at disp01 since the left-hand side of the far wall is currently only represented at disp03. The window will nevertheless trap filled-in signals within the monocular FCS filling-in domains associated with disp01. (e) Finally the picture frame represents a completed structure within the zero disparity disp00 bipole cell pool.

transform the multiple-scale BCS computations into multiple-depth BCS computations. This transformation enables each BCS copy to use all the available evidence to compute the most accurately positioned boundaries possible within its depth range. The resultant depth-sensitive BCS copies control the filling-in of FCS surfaces at the corresponding relative depths from the observer.

The bipole cells implemented here refine the traditional bipole. Firstly, they receive input from multiple spatial scales. In the current simulation, each pool of bipole cells, except that which codes the largest disparity (disp14), sums inputs from both spatial scales. In addition, perpendicular hypercomplex cells inhibit each bipole cell receptive field. This cross-orientation competition realizes the property of *spatial impenetrability* (Grossberg, 1987a; Grossberg & Mingolla, 1987) that helps to prevent boundaries from being colinearly completed over regions that contain mutually perpendicular boundaries, or other nonparallel orientations.

Another innovation is direct input from hypercomplex cells at the location of the bipole cell itself. This term allows bipole cells that receive input from their own location and at least one lobe to trigger outputs, while still preventing boundaries from completing outwards. As a result, the ends of lines can receive feedback from the bipole cells. Gove et al. (1995) and Grossberg et al. (1995) simulate related uses of this refined bipole cell.

Figure 11(a)–(e) shows the open loop response of the bipole cell stage for both the horizontal and vertically tuned cells of disp4, disp3, disp2, disp1, and disp0, respectively. These profiles represent BCS activity before the SOCC loop or monocular FIDOs provide feedback.

Each bipole cell, except those which code the largest disparity (disp14), receives input from both spatial scales as follows:

$$\frac{dz_{ijkd}}{dt} = -z_{ijkd} + h[g(A^b_{s1} + A^b_{s2}) + g(B^b_{s1} + B^b_{s2}) + N_{ijkd}] \tag{36}$$

In eqn (36), $g(x)$ bounds each lobe's activity as follows:

$$g(x) = \frac{[x]^+}{D + [x]^+} \tag{37}$$

whereas the threshold Γ in $h[x]$ ensures that both lobes must be active before the bipole cell fires; namely,

$$h[x] = [x - \Gamma]^+ \tag{38}$$

Terms A^b_{s1}, A^b_{s2} and B^b_{s1}, B^b_{s2} in eqn (36) correspond to the summed inputs from each bipole cell's two lobes for each spatial scale. They are defined generically as follows:

$$A^b = \sum_{(p,q,r)} ((N_{i+p,j+q,rd} - N_{i+p,j+q,Rd})[Z_{ijkr}]^+) \tag{39}$$

and

$$B^b = \sum_{(p,q,r)} ((N_{i+p,j+q,rd} - N_{i+p,j+q,Rd})[-Z_{ijkr}]^+) \tag{40}$$

where $-P^b_1 \le p, q \le P^b_1$ (scale 1), and $-P^b_2 \le p, q \le P^b_2$ (scale 2).

In eqn (36), z_{ijkd} is the bipole cell activity at position ij, orientation k, and disparity d. In eqn (37), D is a parameter of the lobe activity squashing function (0.1). In eqn (38), Γ is the firing threshold of the bipole cells ($d \in \{0, 1, 2, 3\}, \Gamma = 1; d = 4, \Gamma = .5\}$. In eqns (39) and (40), N_{ijkd} is the activity of the hypercomplex cells in eqn (35) and orientation R is perpendicular to orientation r. Terms Z_{ijkr} and $-Z_{ijkr}$ define the lobes of the bipole cell receptive field (see below); P^b_1 and P^b_2 define the size of the input fields of the bipole cells for each spatial scale (5, 5); $r \in \{0, 1\}$ determines the orientation tuning of the cell; and $[x]^+$ denotes half-wave rectification. The central input term N_{ijkd} in eqns (39) and (40) ensures that the ends of lines also strongly activate bipole cells.

In the current implementation, the lobes of the bipole cell receptive field obey the Gaussian weighting operator used by Gove et al. (1995):

where
$$Z_{ijkr} = \text{Zsgn}[i]\exp[T_g + T_k + T_r] \tag{41}$$

$$T_g = \frac{-(\sqrt{i_2 + j_2} + \rho)^2}{2\sigma_g^2} \tag{42}$$

$$T_k = \frac{\left(\frac{k\pi}{T} - \tan^{-1}(i/j)\right)^2}{2\sigma_k^2}$$

and

$$T_r = \frac{\left(\frac{r\pi}{T} - \tan^{-1}(i/j)\right)^2}{2\sigma_r^2} \tag{44}$$

In eqns (41)–(44), Z is a scale constant (1); ρ determines the peak of the Gaussian term T_g (-2.5); σ_g, σ_k and σ_r are the standard deviations of the T_g, T_k and T_r Gaussian terms (1, 0.15, 0.15); and T determines the shape of the lobes (2).

Bipole cells can support boundary grouping and completion within a band of orientations, some of which receive weaker support from image contrasts than others. Competitive interactions help to remove this uncertainty. Orientational and spatial competition are applied to the outputs of the bipole cells as they are fed back to the first hypercomplex cell stage. This feedback pathway mirrors the competitive stages of the feedforward path. Positive feedback from the bipole cells goes back to hypercomplex cells that code the same disparity and position, across all scales (see Figure 2, pathways 4). These feedback operations are defined as follows.

Cross-orientation competition is achieved using a membrane equation:

$$\frac{du_{ijkd}}{dt} = -\alpha_6 u_{ijkd} + (U_6 - u_{ijkd})C_6 - (u_{ijkd} + L_6)E_6 \tag{45}$$

where

$$C_6 = \sum_r C_{kr} z_{ijrd} \text{ with } C_{kr} = \frac{c}{2\Pi\sigma_c^2}\exp\left[-\frac{1}{2}\left(\frac{r-k}{\sigma_c}\right)^2\right] \tag{46}$$

and

$$E_6 = \sum_r E_{kr} z_{ijrd} \text{ with } E_{kr} = \frac{s}{2\Pi\sigma_s^2}\exp\left[-\frac{1}{2}\left(\frac{r-k}{\sigma_s}\right)^2\right] \tag{47}$$

In eqn (45), u_{ijkd} is the activity of the hypercomplex cell at position ij, orientation k and disparity d; α_6 is the decay parameter (1); and U_6 and $-L_6$ are the upper and lower bounds of hypercomplex cell activity (1, -1). In eqns (46) and (47), z_{ijkd} is the bipole cell activity at location ij, orientation k and disparity d; c and s are scale constants (1, 1.5); σ_c and σ_s are the standard deviations of the center and surround Gaussians (scale 1: 0.5, 0.75; scale 2: 0.5, 0.75); and r and k denote orientation ($r \in \{0, 1\}, k \in \{0, 1\}$). Solving at equilibrium and half-wave rectifying yields the output:

$$U_{ijkd} = \frac{\left[\sum_r (U_6 C_{kr} - L_6 E_{kr}) z_{ijkd}\right]^+}{\alpha_6 + \sum_r (C_{kr} + E_{kr}) z_{ijkd}} \tag{48}$$

Spatial competition then occurs in a center-surround manner between cells coding the same relative depth and orientation. Hence

$$\frac{dO_{ijkd}}{dt} = -\alpha_7 O_{ijkd} + (U_7 - O_{ijkd})C_7 - (O_{ijkd} + L_7)E_7 \tag{49}$$

where C_7 and S_7 are defined as follows:

$$C_7 = \sum_{(h,v)} C_{hv} U_{i+h,j+v,k,d}$$

$$C_{hv} = A\exp\left[-\frac{1}{2}\left(\frac{h^2+v^2}{\sigma_c^2}\right)\right] \tag{50}$$

$$E_7 = \sum_{(h,v)} E_{hv} U_{i+h,j+v,k,d}$$

$$E_{hv} = B\exp\left[-\frac{1}{2}\left(\frac{h^2+v^2}{\sigma_s^2}\right)\right] \tag{51}$$

where $-P_1^s \le h,\ v \le P_1^s$ (scale 1), and $-P_2^s \le h, \le P_2^s$ (scale 2).

In eqn (49), O_{ijkd} is the activity of the hypercomplex cell at position ij, orientation k and disparity d; α_7 is the decay parameter (0.1); and U_7 and $-L_7$ are the upper and lower bounds of hypercomplex cell activity (1, -1). In eqns (50) and (51), U_{ijkd} is the hypercomplex cell activity from the previous feedback stage at location ij, orientation k and disparity d; P_1^s and P_2^s define the size of the spatial interaction for scale 1 and scale 2, respectively (4, 4); A and B are scale constants (1, 1); and σ_c and σ_5 are the standard deviations of the center and surround Gaussians (0.15, 1). Solving at equilibrium and half-wave rectifying yields the output

$$O_{ijkd} = \frac{\left[\sum_{(h,v)} (U_7 C_{hv} - L_7 E_{hv}) U_{i+h,j+v,k,d}\right]^+}{\alpha_7 + \sum_{(h,v)} (C_{hv} + E_{hv}) U_{i+h,j+v,k,d}} \tag{52}$$

4.7. Output of the BCS

After several (typically no more than three) iterations through the BCS feedback pathways, the activities of the SOCC loop cells converge. The boundary segmentations that are created in this way have no color or brightness qualities, since the complex cells pool signals from simple cells of opposite contrast polarity. The filling-in process within the FCS creates visible surface percepts, including the visible percepts of single dots and lines, which are processed as small surface regions. The BCS organizes these FCS filling-in events by topographically interacting with the FCS filling-in domains, or FIDOs, as in Figure 2. The BCS → FCS signals play two roles. (1) They activate filling-in within FIDOS whose FCS inputs are topographically aligned with a BCS boundary, and suppress other FCS inputs. In this capacity, they are

called *filling-in generators*. (2) They act as barriers to filling-in and thereby create surface regions within which filling-in is restricted. In this capacity, they are called *filling-in barriers* (Grossberg, 1987b, 1994).

4.8. Monocular FIDOS: Surface Capture and Filling-In

The projections from the binocular BCS segmentations to the monocular FIDOs generate filling-in (see Figure 2, pathway 6) by selecting those monocular FCS signals (transmitted along pathways 5 in Figure 2) that are spatially coincident with the binocular BCS boundaries, while suppressing those that are not. They thereby selectively capture FCS brightness or color signals at different depth planes to initiate depthful surface completion. Boundary outputs also create resistive barriers to this diffusive process. Filling-in obeys the following equations:

$$\frac{dF^+_{ijd}}{dt} = -M^m F^+_{ijd} + \sum_{(p,q)\,\epsilon\, N} (F^+_{pqd} - F^+_{ijd})\Psi^m_{pqijd} + X^+_{ijd} \tag{53}$$

and

$$\frac{dF^-_{ijd}}{dt} = -M^m F^-_{ijd} + \sum_{(p,q)\,\epsilon\, N} (F^-_{pqd} - F^-_{ijd})\Psi^m_{pqijd} + X^-_{ijd} \tag{54}$$

where diffusion occurs with respect to nearest neighbor cells $N = \{(i, j-1), (i, j+1), (i+1, j), (i-1, j)\}$, and boundaries attenuate the diffusion permeability coefficients as follows:

$$\Psi^m_{pqijd} = \frac{\delta}{\kappa + \epsilon(Z_{pqd} + Z_{ijd})} \text{ with } Z_{ijd} = \sum_k z_{ijkd} \tag{55}$$

In eqns (53)–(55), F^+_{ijd} is the filled-in activity and output signal at position ij and disparity d of the ON FIDO; F^-_{ijd} is the filled-in activity at position ij and disparity d of the OFF FIDO; M^m is a decay (0.1); X^+_{ijd} is the ON channel output from the monocular preprocessing stage; X^-_{ijd} is parameter the OFF channel output from the monocular preprocessing stage; κ is a parameter of the boundary squashing function Ψ^m_{pqijd} (1); δ is a spread-scale parameter (100,000); and ϵ is a blocking scale parameter (1000). At equilibrium, each F_{ijd} was computed as the solution of a set of simultaneous equations:

$$F^+_{ijd} = \frac{X^+_{ijd} + \sum_{(p,q)\in N} F^+_{pqd}\Psi^m_{pqijd}}{M^m + \sum_{(p,q)\in N} \Psi^m_{pqijd}}$$

and

$$F^-_{ijd} = \frac{X^-_{ijd} + \sum_{(p,q)\in N} F^-_{pqd}\Psi^m_{pqijd}}{M^m + \sum_{(p,q)\in N} \Psi^m_{pqijd}} \tag{56}$$

These simultaneous equations were solved at equilibrium using the Y12M matrix inversion package of Zlatev et al. (1981).

4.9. Output Signals from the Monocular FIDOS

Output signals from FCS to BCS are generated only from successfully filled-in regions within the monocular FIDOS, since only these regions can generate contrast-sensitive output signals. An on-center off-surround network suppresses uniformly filled-in regions so that only spatial contrasts within each FIDO are passed on for further processing:

$$\frac{dr^+_{ijd}}{dt} = -\alpha_8 r^+_{ijd} + (U_8 - r^+_{ijd})C^+_8 - (r^+_{ijd} + L_8)E^+_8 \tag{57}$$

and

$$\frac{dr^-_{ijd}}{dt} = -\alpha_8 r^-_{ijd} + (U_8 - r^-_{ijd})C^-_8 - (r^-_{ijd} + L_8)E^-_8 \tag{58}$$

where

$$-P < p,\ q < P,$$

$$C^\pm_8 = \sum_{(p,q)} C_{pq} F^\pm_{i+p,j+q,d}$$

with

$$C_{pq} = C(2\pi\sigma_c^2)^{-1}\exp\left\{-\frac{1}{2}((p^2+q^2)/\sigma_c^2)\right\}$$

and

$$E^\pm_8 = \sum_{(p,q)} E_{pq} F^\pm_{i+p,j+q,d} \tag{61}$$

with

$$E_{pq} = E(2\pi\sigma_s^2)^{-1}\exp\left\{-\frac{1}{2}((p^2+q^2)/\sigma_s^2)\right\} \tag{62}$$

In eqns (57) and (58), r^+_{ijd} and r^-_{ijd} are the activity of ON and OFF monocular FIDO cells at position ij and disparity d; α_8 is a decay parameter (0.01); and U_8 and L_8 are upper and lower bounds on the FCS cells activity (1, 1). In eqns (59)–(62), F_{ijd} is the output (56) of the corresponding monocular FIDO filling-in stage; C and E are scale parameters (0.0398, 0.0181); P determines the size of the spatial interaction (7); and σ_c and σ_s are the standard deviations for the center and surround Gaussians (2, 3). At equilibrium, the ON and OFF activities converge to

$$r^+_{ijd} = \frac{\sum_{(p,q)} U_8 C_{pq} F^+_{i+p,j+q,d} - L_8 E_{pq} F^+_{i+p,j+q,d}}{\alpha_8 + \sum_{(p,q)} (C_{pq} + E_{pq}) F^+_{i+p,j+q,d}} \tag{63}$$

and

$$r^-_{ijd} = \frac{\sum_{(p,q)} U_8 C_{pq} F^-_{i+p,j+q,d} - L_8 E_{pq} F^-_{i+p,j+q,d}}{\alpha_8 + \sum_{(p,q)} (C_{pq} + E_{pq}) F^-_{i+p,j+q,d}} \tag{64}$$

These monocular FIDO outputs were then subtracted and rectified to generate double opponent output signals:

$$R_{ijd}^{+} = [r_{ijd}^{+} - r_{ijd}^{-}]^{+} \text{ and } R_{ijd}^{-} = [r_{ijd}^{-} - r_{ijd}^{+}]^{+} \quad (65)$$

The monocular FIDO output signals act by inhibiting the hypercomplex cell activities of more distant boundaries in eqn (32). By reorganizing boundaries, and the surfaces that they contain, in a near-to-far manner, FCS-to-BCS feedback shows how filling-in of surface properties can also occur in a near-to-far manner. In the current implementation, the nearest depth plane was filled in first. Any negative feedback arising from the monocular FIDOs was fed back into the BCS representations of more distant surfaces. Each SOCC loop was again allowed to converge in response to the new boundary configuration before filling-in of the next-to-nearest depth plane occurred. This process was continued until all the depth representations equilibrated.

In order to map the unoriented FIDO activities onto oriented hypercomplex cell activities, they were first processed by oriented filters. The total signal from each position ij, orientation k and disparity d is defined by:

$$v_{ijkd} = b_{ijkd}^{\text{odd}+} + b_{ijkd}^{\text{odd}-} + b_{ijkd}^{\text{even}+} + b_{ijkd}^{\text{even}-} \quad (66)$$

where

$$b_{ijkd}^{\text{odd}+/-} = \left[\sum_{(p,q)} B_{pqk}^{\text{odd}} R_{i-p,j-q,d}^{+/-} - \sum_{(p,q)} B_{pqk}^{\text{odd}} R_{i+p,j+q,d}^{-/+} \right]^{+} \quad (67)$$

$$b_{ijkd}^{\text{even}+/-} = \left[\sum_{(p,q)} B_{pqk}^{\text{even}} R_{i-p,j-q,d}^{+/-} - \sum_{(p,q)} B_{pqk}^{\text{even}} R_{i+p,j+q,d}^{-/+} \right]^{+} \quad (68)$$

$$B_{pqk}^{\text{odd}} = A_1 \sin\left(\frac{2\Pi K}{T}\right) \exp\left[-\frac{1}{2}\left(\frac{p^2}{\sigma_{pk}^2} + \frac{q^2}{\sigma_{qk}^2}\right)\right] \quad (69)$$

and

$$B_{pqk}^{\text{even}} = A_2 \cos\left(\frac{2\Pi K}{T}\right) \exp\left[-\frac{1}{2}\left(\frac{p^2}{\sigma_{pk}^2} + \frac{q^2}{\sigma_{qk}^2}\right)\right] \quad (70)$$

with $-S_1 \le p, q \le S_1$ (scale 1), and $-S_2 \le p, q \le S_2$ (scale 2).

In eqns (67) and (68), R_{ijd}^{+} is the FCS ON cell response at position ij and disparity d, and R_{ijd}^{-} is the FCS OFF cell response; see eqn (65). In eqns (69) and (70), A_1 and A_2 are scale constants (1, 1); T determines the periodicity of the cell (scale 1: even $T = \pi$, odd $= 2\pi$; scale 2: even $= \pi$, odd $= 3\pi$); S_1 and S_2 define the size of the hypercomplex cells input field (4, 6); σ_{pk} and σ_{qk} define the standard deviations of the x- and y-dimensions of the hypercomplex cell input field for orientation k (scale 1: vertical: $\sigma_p = 2$, $\sigma_q = 1.5$; horizontal: $\sigma_p = 1.5$, $\sigma_q = 2$. Scale 2: vertical: $\sigma_p = 2.5$, $\sigma_q = 2$; horizontal: $\sigma_p = 2$, $\sigma_q = 2.5$).

Figure 12(a)–(d) show the simulated output of the monocular ON and OFF left and right eye FIDOs in response to the da Vinci stereopsis display. ON and OFF activities are graphed together in each figure. No activity is registered in the FIDOs corresponding to disp04. Figure 12(a) displays the ON and OFF activities associated with the near surface which is registered in the FIDOs corresponding to disp03. The left eye's representation differs slightly from that of the right. This is due to a difference in the filling-in signal strength associated with the left edge of the window. Figure 12(b) displays the monocular FCS outputs for depth plane 2. Although no connected figures were present within the BCS disp02 representation, a small filled-in signal is trapped by one or two of the nonconnected boundaries. This spurious noise, which is dependent on the parameters of the FCS filling-in process, does not cause any subsequent problems for the FACADE segmentation. Figure 12(c) displays the monocular FIDO outputs for the far wall and window. Although the far wall does not form a connected figure within the monocular FCS, ON and OFF signals are registered for the right-hand wall. This is a property of the FCS flow parameters and is eliminated by the subsequent stages of the FIDO processing. Finally, Figure 12(d) displays the ON and OFF FIDO representation of the background. A small amount of noise is also seen in both displays.

4.10. The Binocular FIDOS: FACADE Representation

The binocular FIDOs represent the final level of surface processing in the FACADE model, as in Figure 2. Their filled-in representations multiplex properties of surface form, color, depth, orientation, position, and brightness. Grossberg (1994) has compared these receptive field profiles with analogous neurophysiological data from extrastriate area V4 (Zeki, 1983a, 1983b; Desimone et al., 1985). In the da Vinci stereogram percept, the monocularly viewed region BC in Figure 1 has its surface properties filled-in within the same depth-selective FIDO as the binocularly fused boundaries of region CD.

Figure 13(a)–(c) displays the final output of the FACADE model. Figure 13(a) displays the FACADE representation of the near surface which is correctly filled in within the FIDOs that correspond to disparity 3. Figure 13(b) displays the FACADE representation of the filled-in surfaces at disparity 1. Both the window and the far wall are represented at this disparity. Finally Figure 13(c) represents the background which is correctly rendered as a filled-in surface at zero disparity. The FCS representations associated with disparities 2 and 4 contain ganzfelds of activity as no connected regions exist at these depths.

Several factors make this result possible. Perhaps the most important one is boundary enrichment. If the

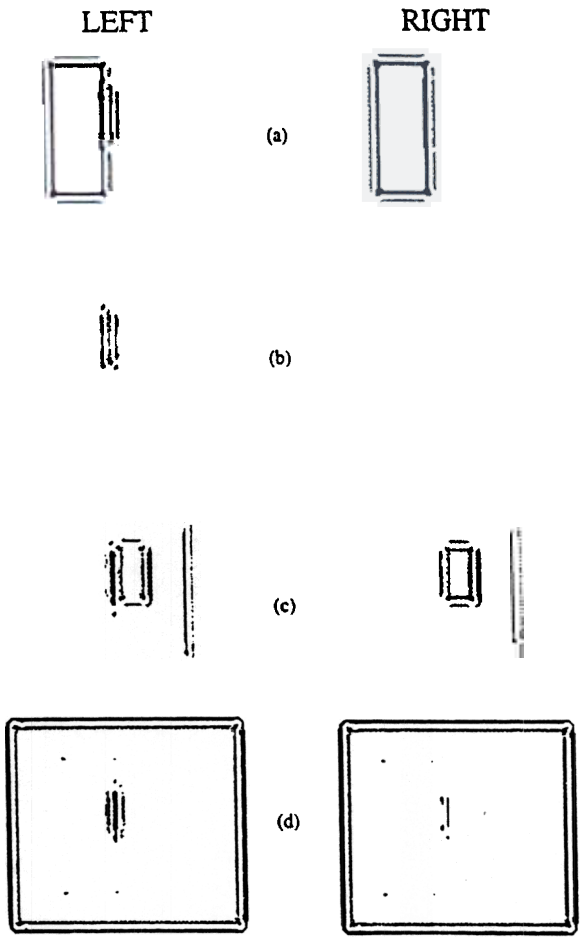

FIGURE 12. Outputs from the monocular ON and OFF left and right eye FIDOs. Disp04 is not represented as no output signals were generated at that disparity. (a) The near surface (AB) is represented within the outputs of the disp03 FIDOs. (b) Some spurious noise occurs within the outputs of the disp02 FIDOs. This noise is caused when feature signal are trapped within BCS boundary signals. This system noise does not cause any problems for the binocular FIDOs because monocular FCS signals are binocularly matched before visible filling-in occurs. (c) The far window is clearly represented in the outputs of the left and right eye FIDOs. We can also see some feature signal trapped by the right-hand side of the far wall. (d) The window frame is correctly represented in the outputs of the ON and OFF monocular FIDOs. Again a small amount of noise is registered within the outputs of the disp00 FIDOs.

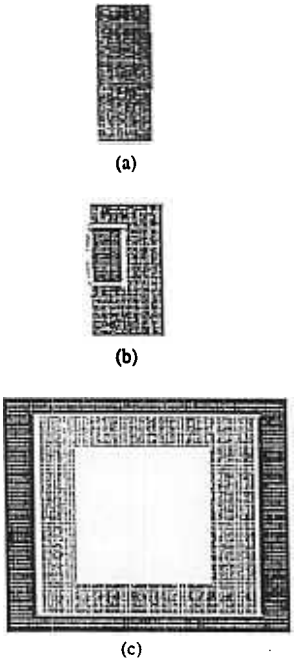

FIGURE 13. Final outputs from the FACADE model.The disp04 and disp02 representations are not displayed. Since no connected regions exist within these depth planes, the disp04 and disp02 representations resemble ganzfelds. (a) The near surface AB is filled-in with its surface properties at disp03. (b) The far wall and window is filled-in within the disp01 syncytia. (c) The picture frame and background are filled-in with their surface properties at disparity 0 (disp00).

binocular boundary at B were processed only as part of the connected region AB, then, as in Figure 14(a), there would be nothing to stop the surface properties of region BD from filling-in behind AB, since BD does not possess a connected boundary. FACADE theory remedies this problem by asserting that the boundaries corresponding to nearer objects are added to the boundaries corresponding to farther objects at the binocular FIDOs. As a result, region BD is surrounded by a connected boundary and its surface properties do not spill out, as shown in Figure 14(b). This restriction upon FCS filling-in does not, however, prevent BCS boundaries from being amodally completed behind occluding figures (Grossberg, 1994).

The binocular FIDOs receive their FCS inputs from both the left and right eye monocular FIDOs and the monocular preprocessing stage. Input from the preprocessing stage is excitatory, while surface pruning inhibitory signals are produced by the monocular FIDOs. Both excitatory and inhibitory signals are binocularly matched via the following membrane equations before they trigger filling-in:

$$\frac{da_{ijd}^{+}}{dt} = -\alpha_9 a_{ijd}^{+} + (U_9 - a_{ijd}^{+})E_{ijd}^{+} - (L_9 + a_{ijd}^{+})I_{ijd}^{+} \quad (71)$$

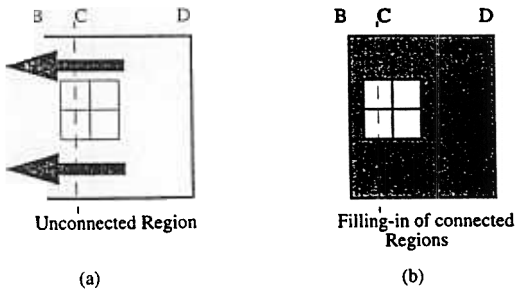

FIGURE 14. Filling-in produces visible percepts only within connected regions. (a) When a region is not fully connected, the featural signal can flow out. (b) Only a fully connected region can trap the FCS signal and produce a visible percept.

and

$$\frac{da_{ijd}^{-}}{dt} = -\alpha_9 a_{ijd}^{-} + (U_9 - a_{ijd}^{-})E_{ijd}^{-} - (L_9 + a_{ijd}^{-})I_{ijd}^{-} \quad (72)$$

$$E_{ijd}^{+} = X_{ijd}^{L+} + X_{ijd}^{R+}, \; E_{ijd}^{-} = X_{ijd}^{L-} + X_{ijd}^{R-} \quad (73)$$

$$I_{ijd}^{+} = \sum_{e>d} R_{ije}^{+}, \text{ and } I_{ije}^{-} = \sum_{e>d} R_{ije}^{-} \quad (74)$$

In eqns (71) and (72), a_{ijd}^{+} and a_{ijd}^{-} net inputs to the ON and OFF binocular FIDO cells at position ij and disparity d; α_9 is a decay parameter (0.01); and U_9 and $-L_9$ are the upper and lower bounds of activity (1, −1). In eqns (73) and (74), X_{ijd}^{L+} and X_{ijd}^{R+} are the matched monocular inputs (10) from the ON channel monocular preprocessing stage; inputs X_{ijd}^{L-} and X_{ijd}^{R-} come from (11), the OFF channel monocular preprocessing stage; R_{ijd}^{+} is the output (65) of the monocular FIDO ON cell channel at position ij and disparity d; and R_{ijd}^{-} is the corresponding output of the monocular FIDO OFF cell channel. Solving at steady state yields

$$a_{ijd}^{+} = \frac{U_9 E_{ijd}^{+} - L_9 I_{ijd}^{+}}{\alpha_9 + E_{ijd}^{+} + I_{ijd}^{+}} \text{ and } a_{ijd}^{-} = \frac{U_9 E_{ijd}^{-} - L_9 I_{ijd}^{-}}{\alpha_9 + E_{ijd}^{-} + I_{ijd}^{-}} \quad (75)$$

These signals trigger filling-in of binocular FIDOs just as in the monocular FCS filling-in process [eqns (53) and (54)] stage. The diffusive spread of activity is defined by the following equations:

$$\frac{d\Omega_{ijd}^{+}}{dt} = -M^b \Omega_{ijd}^{+} + \sum_{(p,q)\in N} (\Omega_{pqd}^{+} - \Omega_{ijd}^{+}) \Psi_{pqijd}^{b} + a_{ijd}^{+} \quad (76)$$

$$\frac{d\Omega_{ijd}^{-}}{dt} = -M^b \Omega_{ijd}^{-} + \sum_{(p,q)\Psi N} (\Omega_{pqd}^{-} - \Omega_{ijd}^{-}) \Psi_{pqijd}^{b} + a_{ijd}^{-} \quad (77)$$

where $N = \{(i,j-1),(i,j+1),(i+1,j),(i-\ ,j)\}$, and

$$\Psi_{pqijd}^{b} = \frac{\delta}{\kappa + \epsilon(Z_{pqd} + Z_{ijd})} \text{ with } Z_{ijd} = \sum_{k,e\geq d} z_{ijke} \quad (78)$$

In eqns (76)–(78), Ω_{ijd}^{+} is the filled-in value at position ij and disparity d of the ON FIDO; Ω_{ijd}^{-} is the corresponding OFF value; M^b is a decay parameter (0.1); a_{ijd}^{+} in eqn (76) is the ON channel filling-in signal; a_{ijd}^{+} in eqn (77) is the OFF channel signal; κ in eqn (78) is a parameter of the boundary squashing function Ψ_{pqijd}^{b} (1), δ is a spread scale parameter (10,000), and ϵ is a blocking scale parameter (100).

At equilibrium, each Ω_{ij} is computed as the solution of a set of simultaneous equations:

$$\Omega_{ijd}^{+} = \frac{a_{ijd}^{+} + \sum_{(p,q)\in N} \Omega_{pqd}^{+} \Psi_{pqijd}^{b}}{M^b + \sum_{(p,q)\in N} \Psi_{pqijd}^{b}}$$

and

$$\Omega_{ijd}^{-} = \frac{a_{ijd}^{-} + \sum_{(p,q)\in N} \Omega_{pqd}^{-} \Psi_{pqijd}^{b}}{M^b + \sum_{(p,q)\in N} \Psi_{pqijd}^{b}}$$

These simultaneous equations were also solved using the Y12M matrix inversion package.

5. PARAMETER ROBUSTNESS

As in any complex modeling endeavor, the FACADE simulation contains many parameters. The complexity of such a model is evaluated, not in terms of the number of its parameters, but in terms of how many processing stages it includes, what perceptual principles are realized by these stages, how much empirical evidence supports each stage, and how robustly the parameters that define each stage generate that stage's desirable qualitative properties. Grossberg (1994, 1997) summarized perceptual principles and data to argue that each stage is needed. The present refinements further develop these stages. Most of the parameters defining these stages were hand-set to produce realistic looking receptive field profiles; for example, the parameters defining the ON and OFF center-surround receptive fields. A wide variety of parameter choices reproduced similar responses. A decision to implement only near (positive) disparity sensitive complex cells, along with zero disparity and monocular complex cells, was taken as the simulated scene contains only near surfaces when the model fixates the image frame. This decision, along with the choice of the explicit disparities that were coded, set our binocular combination parameters.

The binocular filter is perhaps the most parameter-sensitive aspect of the model; however, an appropriate set of matching parameters can be learned (Grunewald & Grossberg, 1997). The parameters involved in the final stages of the model are also valid within wide ranges of parameters. Important parameters involved in this process include T, the tonic activity of the hypercomplex cells in eqn (28), which must be set to be greater than zero for the endstopping process to have its desired effect, and Γ in eqn (38), which if incorrectly set can allow contours to complete outward from line ends in an unrealistic manner. The parameters chosen in both the monocular and binocular FCS simulations were also robust. In fact, as Figure 12 shows, this choice of parameters allowed spurious BCS activity to trap minor quantities of FCS signal. As the binocular FIDOs remove this noise, we did not deem it necessary to search for more ideal parameter settings. A wide range of values exist that produce comparable responses. If, however, parameters δ, κ and ϵ from eqns (55) and (78) are set to extreme values, for example to zero, filling-in can be stopped altogether or allowed to pass through BCS boundaries without any obstruction.

6. GENERAL DISCUSSION

The present article develops the first rigorous implementation of the three-dimensional FACADE model and shows how it can combine binocularly fused and monocularly unfused boundaries within the BCS to generate a three-dimensional surface representation of the visual scene within the FCS. The addition of boundaries from monocular to binocular cell pools suggests how monocularly viewed regions might be incorporated into a global three-dimensional percept. However, if this is the case, then why do we not see double images corresponding to the monocular representations of binocularly fused boundaries? Our answer is that the activities within the monocular, zero disparity, and nonzero-disparity cell pools that are used to form the final percept are emergent properties of the boundary segmentation process. Cell pools interact in a competitive manner to ensure that false or weak matches are suppressed whenever possible. This inhibition is assumed to be between cells that attempt to code the same inputs and is tuned developmentally such that cells that fire together (or to the same input pattern) wire together (in an inhibitory manner).

The model includes a new multiscale simple-to-complex cell filter that reconciles the need to binocularly match left and right eye images of the same contrast polarity with the equally important need to form boundaries around objects whose relative contrasts with their backgrounds reverse along their perimeter. The model also notes how pooling inputs from even and odd symmetric simple cells at complex cells eliminates some spurious binocular matches.

Perhaps the most important general contribution of the model is its rigorous exposition of how early mechanisms of stereo matching interact with the later mechanisms of three-dimensional boundary segmentation and surface representation that lead to visual percepts. This link clarifies how percepts of occluding and occluded objects may form in response to both three-dimensional scenes and two-dimensional pictures, and how visual illusions may be generated by the same mechanisms that produce our emergent percepts of the real world.

REFERENCES

Anderson, B.L., & Nakayama, K. (1994). Toward a general theory of stereopsis: binocular matching, occluding contours, and fusion. *Psychological Review, 101,* 414–445.

Belhumeur, P.N., & Mumford, D. (1992). A Bayesian treatment of the stereo correspondence problem using half-occluded regions. *Proceedings of the Institute of Electrical and Electronics Engineers Conference on Computer Vision and Pattern Recognition* (pp. 506–512). Los Alamitos, CA: IEEE.

Brewster, D. (1856). *The Stereoscope. Its history, theory, and construction.* London: John Murray.

Chubb, C., & Sperling, G. (1989). Two motion perception mechanisms revealed through distance-driven reversal of apparent motion. *Proceedings of the National Academy of Science U.S.A., 86,* 2985–2989.

Cohen, M.A., & Grossberg, S. (1984). Neural dynamics of brightness perception: features, boundaries, diffusion, and resonance. *Perception and Psychophysics, 36,* 562–570.

Desimone, R., Schein, S.J., Moran, J., & Ungerleider, L.G. (1985). Contour, color, and shape analysis beyond the striate cortex. *Vision Research, 25,* 441–452.

de Yoe, E.A., & van Essen, D.C. (1988). Concurrent processing streams in monkey visual cortex. *Trends in Neuroscience, 11,* 219–226.

Emmert, E. (1881). Grössenverhaltnisse der Nachbilder. *Klinische Monatsblätter für Augenheilkunde, 19,* 443–450.

Euclid (1557). *Euclidis Optica and catoptrica* (Andreae Wecheli, Trans.) Paris.

Frisby, J.P., & Pollard, S.B. (1991). Computational issues in solving the stereo correspondence problem. In M.S. Landy and J.A. Movshon (Eds.), *Computational models of visual processing.* Cambridge, MA: MIT Press.

Galen (1550/1968). *De usu partium corporis humani* (M.T. May, Trans.). New York: Cornell University Press.

Gogel, W.C. (1956). The tendency to see objects as equidistant and its reverse relations to lateral separation. *Psychological monograph, 70,* (Whole no. 411).

Gogel, W.C. (1965). Equidistance tendency and its consequences. *Psychological Bulletin, 64,* 153–163.

Gogel, W.C. (1970). The adjacency principle and three-dimensional visual illusions. *Psychonomic Monograph Supplement, 3,* 153–169.

Gonzalez, F., Revola, J.L., Perez, R., Acufia, C., & Alonso, J. (1993). Cell responses to vertical and horizontal retinal disparities in the monkey visual cortex. *Neuroscience Letters, 160,* 167–170.

Gove, A., Grossberg, S., & Mingolla, E. (1995). Brightness perception, illusory contours, and corticogeniculate feedback. *Visual Neuroscience, 12,* 1027–1052.

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

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

Grossberg, S. (1987a). Cortical dynamics of three-dimensional form, color, and brightness perception: I. Monocular theory. *Perception and Psychophysics, 41,* 87–116.

Grossberg, S. (1987b). Cortical dynamics of three-dimensional form, color, and brightness perception: II. Binocular theory. *Perception and Psychophysics, 41,* 117–138.

Grossberg, S. (1993). Boundary, brightness, and depth interactions during preattentive representation and attentive recognition of figure and ground. *Italian Journal of Psychology, 20,* 771–804.

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

Grossberg, S. (1997) *Cortical dynamics of 3-D figure-ground perception of 2-D pictures. Psychological Review, 104,* 618–658.

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., & Mingolla, E. (1985). Neural dynamics of perceptual grouping: textures, boundaries, and emergent segmentations. *Perception and Psychophysics, 38,* 141–171.

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

Grossberg, S., Mingolla, E., & Williamson, J. (1995). Synthetic aperture radar processing by a multiple scale neural system for boundary and surface representation. *Neural Networks, 8,* 1005–1028.

Grossberg, S., & Todorovic, 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.

Grunewald, A., & Grossberg, S. (1997). Self-organization of binocular disparity tuning by reciprocal corticogeniculate interactions. Tech Report CAS/CNS-TR-96-027, Boston, MA: Boston University. *Journal of Cognitive Neuroscience* (in press).

Harris, J., & Parker, A. (1995). Independent neural mechanisms for bright and dark information in binocular stereopsis. *Nature, 374,* 808–811.

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

Hubel, D.H., & Wiesel, T.N. (1965). Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat. *Journal of Neurophysiology, 28,* 229–289.

Julesz, B., & Miller, J.E. (1975). Independent spatial frequency tuned channels in binocular vision and rivalry. *Perception, 4,* 125–143.

Julesz, B., & Schumer, R.A. (1978). Early visual perception. *Annual Review of Psychology, 32,* 572–627.

Katz, L.C., & Callaway, E.M. (1992). Development of local circuits in mammalian visual cortex. *Annual Review of Neuroscience, 15,* 31–56.

Kaufman, L. (1974). *Sight and mind.* New York: Oxford University Press.

Kulikowski, J.J. (1978). Limits on single vision in stereopsis depends on contour sharpness. *Nature, 275,* 126–127.

LeVay, S., & Voight, T. (1988). Ocular dominance and disparity coding in cat visual cortex. *Visual Neuroscience, 1,* 395–414.

Liu, L., Tyler, C., & Schor, C. (1992). Failure of rivalry at low contrast: evidence of a suprathreshold binocular summation process. *Vision Research, 32,* 1471–1479.

Löwel, S., & Singer, W. (1992). Selection of intrinsic horizontal connections in the visual cortex by correlated neuronal activity. *Science, 255,* 209–212.

Marr, D. (1980). *Vision.* New York: W.H. Freeman and Company.

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

Marr, D., & Poggio, T. (1979). A computational theory of human stereo vision. *Proceedings of the Royal Society of London (B), 204,* 301–328.

McKee, S.P. (1993). Stereomatching and binocular contrast summation. *Perception,* 22, ECVP (Suppl. 29).

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

McKee, S.P., & Harrad, R. (1993). Fusional suppression in normal and stereoanomalous observers. *Vision Research, 33,* 1645–1658.

McKee, S.P., Levi, D., & Bowne, S. (1990). The imprecision of stereopsis. *Vision Research, 30,* 1763–1779.

McLoughlin, N., & Grossberg, S. (1994). How are monocularly viewed surfaces perceived in depth during da Vinci stereopsis? *Investigative Ophthalmology and Visual Science, 35,* 2110.

McLoughlin, N., & Grossberg, S. (1995). The contrast-sensitive nature of binocular matching: simulations of dichoptic masking and Panum's limiting case. *Investigative Ophthalmology and Visual Science, 36,* 666.

McLoughlin, N., & Grossberg, S. (1997). *Cortical computation of stereo disparity.* Technical Report CASICNS-TR-96-022, Boston, MA: Boston University. *Vision Research* (in press).

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

Orban, G.A., Kato, H., & Bishop, P.O. (1979). Dimensions and properties of end-zone inhibitory areas in receptive fields of hypercomplex cells in cat striate cortex. *Journal of Neurophysiology, 42,* 833–849.

Pessoa, L., Mingolla, E., & Neumann, H. (1995). A contrast- and luminance-driven multiscale network model of brightness perception. *Vision Research, 35,* 2201–2223.

Poggio, G.F. (1984). Processing of stereoscopic information in primate visual cortex. In G.M. Edelman, W.E. Gall, & W. M. Cowan (Eds.), *Dynamical aspects of neocortical function* (pp. 613–635). New York: Wiley.

Poggio, G.F., Gonzalez, E., & Krause, F. (1988). Stereoscopic mechanisms in monkey visual cortex: binocular correlation and disparity selectivity. *Journal of Neuroscience, 8,* 4531–4550.

Pollen, D.A., Gaska, J.P., & Jacobson, L.D. (1989). Physiological constraints on models of visual cortical function. In R.M.J. Cotterill (Ed.), *Models of brain function* (pp. 115–135). Cambridge: Cambridge University Press.

Pollen, D.A., Foster, K.H. & Gaska, J.P. (1985). Phase-dependent response characteristics of visual cortical neurons. In D. Rose, & V.G. Dobson (Eds.), *Models of visual cortex* (pp. 281–291). New York: Wiley.

Pollen, D.A., & Ronner, S.E. (1981). Phase relationships between adjacent simple cells in the visual cortex. *Science, 212,* 1409–1411.

Pollen, D.A., & Ronner, S.E. (1982). Spatial computation performed by simple and complex cells in the visual cortex of the cat. *Vision Research, 22,* 101–118.

Porta, (1593). *De humana physiognomonia libr IIII.* Hanoviae.

Richards, W.A. (1971). Anomalous stereoscopic depth perception. *Journal of the Optical Society of America, 61,* 410–415.

Richards, W.A., & Kaye, M.G. (1974). Local versus global stereopsis: two mechanisms. *Vision Research, 14,* 1345–1347.

Schatz, C.J. (1992). The developing brain. *Scientific American, September,* 61–67.

Schiller, P. (1992). The ON and OFF channels of the visual system. *Trends in Neurosciences, 15,* 86–92.

Schor, C.M., & Tyler, C.W. (1981). Spatio-temporal properties of panum's fusional area. *Vision Research, 21,* 683–692.

Schor, C.M., & Wood, I. (1983). Disparity range for local stereopsis as a function of luminance spatial frequency. *Vision Research, 23,* 1649–1654.

Schor, C.N., Wood, I., & Ogawa, J. (1984). Binocular sensory fusion is limited by spatial resolution. *Vision Research, 24,* 661–665.

Shimojo, S., Bauer, J.A., O'Connell, K.M., & Held, T. (1986). Pre-stereoscopic binocular vision in infants. *Vision Research, 26,* 501–510.

Stevenson, S.B., Cormack, L.K., Schor, C.M., & Tyler, C.W. (1992). Disparity tuning in mechanisms of human stereopsis. *Vision Research, 32,* 1685–1689.

Sutter, A., Beck, J., & Graham, N. (1989). Contrast and spatial variables in texture segregation: testing a simple spatial-frequency channels model. *Perception and Psychophysics, 46,* 312–332.

Tyler, C.W. (1975). Spatial organization of binocular disparity sensitivity. *Vision Research, 15,* 583–590.

Tyler, C.W. (1983). Sensory processing of binocular disparity. In C.M. Schor, & K.J. Cuiffreda (Eds.), *Vergence eye movements* (pp. 199–295). Boston, MA: Butterworths.

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

von Helmholtz, H. (1910/1925). *Treatise on physiological optics* (J. P. Southall, Trans.). New York: Dover.

von Tschermak-Seysenegg, A. (1952). *Introduction to physiological optics* (P. Boeder, Trans.). Springfield, IL: Thomas.

Werner, H. (1937). Dynamics in binocular depth perception. *Psychological monograph* (Whole no. 218).

Zeki, S. (1983a). Colour coding in the cerebal cortex: the reaction of cells in monkey visual cortex to wavelengths and colours. *Neuroscience, 9,* 741–766.

Zeki, S. (1983b). Colour coding in the cerebral cortex: the responses of wavelength-selective and colour coded cells in monkey visual cortex to changes in wavelength composition. *Neuroscience, 9,* 767–791.

Zlatev, Z., Wasniewski, J., & Schaumburg, K. (1981). *YJ2M solution of large and sparse systems.* Berlin: Springer.