FIGURE-GROUND SEPARATION OF CONNECTED SCENIC FIGURES: BOUNDARIES, FILLING-IN, AND OPPONENT PROCESSING

Stephen Grossberg† and Lonce Wyse‡
Center for Adaptive Systems
and
Department of Cognitive and Neural Systems
Boston University
111 Cummington Street
Boston, MA 02215

Abstract

A neural network model is described for automatic parallel separation of connected scenic figures from one another and from their backgrounds. The model is part of a self-organizing architecture for invariant pattern recognition in a cluttered environment. The figure-ground separation process iterates operations from a Feature Contour System (FCS) and a Boundary Contour System (BCS) in the order FBF. The FCS discounts the illuminant and fills-in surface properties, such as brightness and color, using the discounted signals. A key idea of the FBF network is to use filling-in for figure-ground separation. The BCS generates boundary segmentations that define the regions in which filling-in occurs. The BCS is modelled by a feedforward network, call the CORT-X 2 filter, that combines oriented receptive fields with rectifying, competitive, and cooperative interactions to detect, regularize, and complete boundaries in up to 50% analog noise. This filter combines complementary properties of large receptive fields and small receptive fields, and of on-cells and off-cells, to generate positionally more accurate and less noisy boundaries. Double opponent interactions of on-cells and off-cells facilitate separation of figures with incomplete CORT-X boundaries. The results clarify why an FBF network can rapidly separate figures that humans cannot separate during visual search tasks.

1. Theoretical Background: Figure-Ground Separation by Humans and Machines

This chapter contributes to the development of a self-organizing neural network architecture for invariant pattern recognition in a cluttered environment. Carpenter and Grossberg (1988) described a version of this architecture. A related scheme is considered herein (Figure 1), whose primary functional stages are:

† Supported in part by the Air Force Office of Scientific Research (AFOSR 90-0175), the Army Research Office (ARO DAAL-03-88-K0088), DARPA (AFOSR 90-0083), and Hughes Research Laboratories (S1-804481-D and S1-903136).

‡ Supported in part by the American Society for Engineering Education and Hughes Research Laboratories (S1-804481-D).

Acknowledgements: The authors wish to thank Cynthia E. Bradford, Carol Yanakakis Jefferson, and Diana Meyers for their valuable assistance in the preparation of the manuscript.

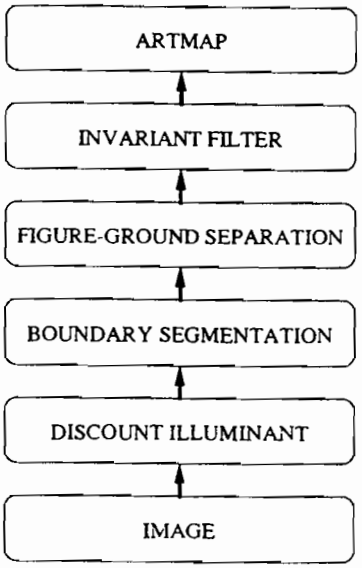

Figure 1. Stages of a neural network architecture for real-time automatic target recognition.

Stage 1. Discount the illuminant.

Stage 2. Detect, regularize, and complete figure boundary. Suppress interior and exterior image noise.

Stage 3. Detach figure from ground.

Stage 4. Filter to give invariance under translation, rotation, and contraction.

Stage 5. Let invariant spectra of the boundary-enhanced, noise-suppressed, detached figures be the input patterns to an ART or ARTMAP architecture for stable self-organization of recognition categories (Carpenter and Grossberg, 1987a, 1987b, 1988; Carpenter, Grossberg, and Reynolds, 1991; Carpenter, Grossberg, Markuzon, Reynolds, and Rosen, 1991, and this volume).

In Carpenter, Grossberg, and Mehanian (1989), a neural network preprocessor for the second stage—the boundary segmentation stage—of such an architecture was described. This boundary preprocessor, called the CORT-X filter, detects, regularizes, and completes sharp (even one pixel wide) boundaries, while simultaneously suppressing the noise. The CORT-X filter is based upon the biologically derived Boundary Contour System of Grossberg and Mingolla (1985a, 1985b, 1987). The Boundary Contour System uses nonlinear feedback interactions to select and complete sharp boundaries over long gaps in image contours. The CORT-X model uses only feedforward interactions that are faster to simulate and easier to implement in hardware. Their ability to complete boundaries is also more limited, but is adequate for many applications. The CORT-X filter is modified herein to deal with at least

50 percent analog noise. This modified filter is called CORT-X 2 to distinguish it from the original model.

We use the CORT-X 2 model to develop a new approach to designing the first three stages of the architecture. We show how Stages 1 and 2 can be designed to facilitate figure-ground separation by Stage 3. Figure-ground separation is the process whereby a figure, or object, in a scene is separated from other figures and background clutter. Whereas knowledge about a figure may facilitate its separation, such knowledge is not necessary for biological vision systems to carry out figure-ground separation. Experiences abound of unfamiliar figures that "pop out" from their backgrounds before they ever enter our corpus of learned knowledge about the world. The fact that figure-ground separation can occur even for unfamiliar figures contributes to the general-purpose nature of biological vision, which can process both unfamiliar and familiar scenes, and does not require prior instruction about an environment in order to operate effectively.

In this chapter, a new type of system is described that is capable of automatic figure-ground separation (Figure 2). This process separates scenic figures whose emergent boundary segmentations (defined below) surround a connected region. As a result of this property, such a system can automatically distinguish between connected and disconnected spirals (Figure 3), a benchmark that gained fame through its emphasis in the book by Minsky and Papert (1969, 1988) on perceptrons. Why the present biologically-motivated algorithm can distinguish interleaved spirals in a way that humans cannot is described below in Section 7. This analysis also clarifies why humans cannot rapidly detect conjunctions of some visual features, such as shape and color, but can rapidly detect conjunctions of other visual features, such as disparity and color, or motion and color (Nakayama and Silverman, 1986; Treisman and Gelade, 1980), that distinguish target figures from surrounding distractors.

Figure-ground separation is an essential step in pattern recognition whenever the objects to be recognized may vary in their position, orientation, and size in a scene. Once a scenic figure has been separated from other figures and background clutter, as in Figure 2, it can be input to an invariant filter at Stage 4 of the architecture (Figure 1). The output of the filter is invariant under translations, rotations, and contractions of the figure. If the figure is not first detached from the scenic background, then output of the filter is not invariant under translations, rotations, and contractions of the figure with respect to the fixed background.

There exist at least three different approaches to automatic figure-ground separation. Section 2 reviews a method for figure-ground separation that uses combinations of laser radars, or related artificial detectors. Sections 3–5 review a method that arises in a neural network model of biological vision that has been called FACADE Theory (Grossberg, 1987, 1990; Grossberg, Mingolla, and Todorović, 1989). Section 6 begins the exposition of the model, which we call an FBF model for reasons that are noted below. The FBF model is capable of separating connected figures from their backgrounds in response to either monochromatic images, such as a grey-scale high-altitude photograph, or from images derived from multiple detectors. Its mechanisms are based upon those described in FACADE Theory, which are adapted for use in a setting where only a single detector may be available.

Figure-ground separation is accomplished in the model by iterating operations adapted from the Feature Contour System (FCS) and the Boundary Contour System (BCS) of FACADE Theory in the order FCS–BCS–FCS; hence the name FBF model. The FCS operations include the use of nonlinear shunting networks to compensate for variable illumi-

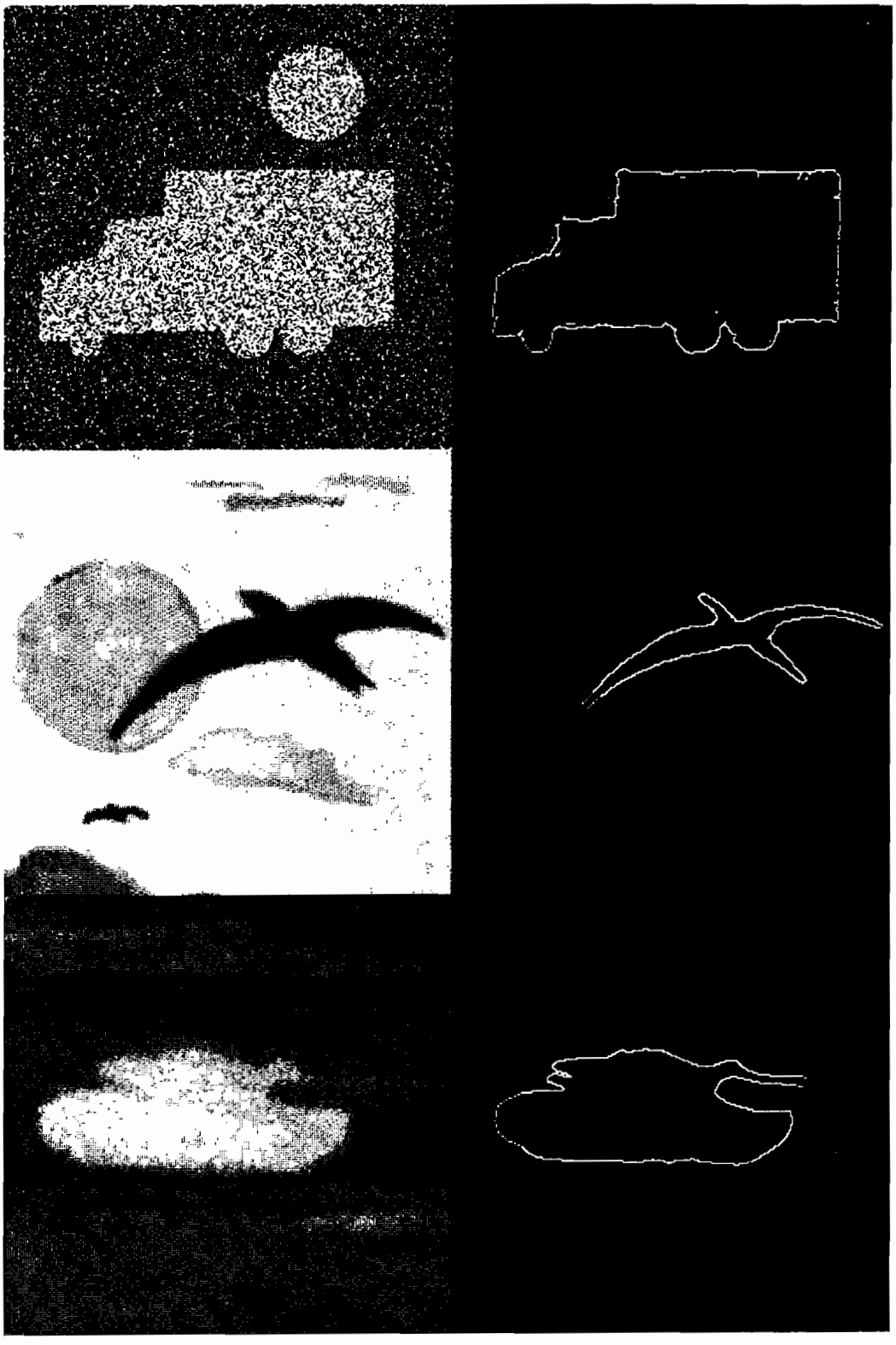

Figure 2. Three examples of figure-ground separation by Stages 1–3 of the architecture.

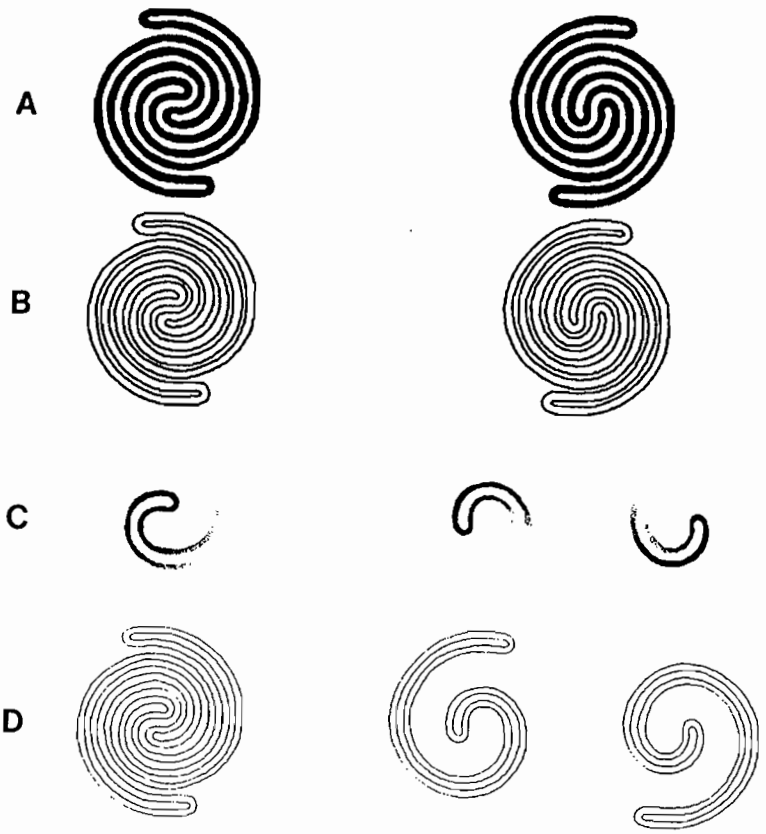

Figure 3. (a) Input images. The left image is connected. The right image consists of two disconnected components. (b) Boundary segmentations. (c) Beginning of filling-in. (d) Separated outputs. The left boundary lies on a single Stage 3 slab. Each connected component of the right boundary lies on a different slab.

nation ("discount the illuminant") and nonlinear diffusion networks to control filling-in. A key new feature of an FBF network is the use of filling-in for figure-ground separation. The BCS operations include oriented filters joined to competitive and cooperative interactions designed to detect, regularize, and complete boundaries in up to 50 percent noise, while suppressing the noise. The new CORT-X 2 filter achieves this competence by using both on-cells and off-cells to generate a boundary segmentation from a noisy image. On-cells and off-cells exhibit two different types of complementary responses that are useful for processing noisy images: they respond in a complementary way to noise pixels (light on dark, vs. dark on light) and to convex and concave image contours. These complementary reactions are joined together to form a boundary segmentation that overcomes the weaknesses of either detector taken separately (Section 6). As a result, the output of such a segmentation process is capable of joining together image features derived from opposite directions-of-contrast. Another new idea is the use of a double-opponent network, defined in Section 6, to facilitate

figure-ground separation of regions with incomplete boundary segmentations.

2. Figure-Ground Separation by Artificial Detectors: Laser Radar Arrays

A conceptually simple technique for figure-ground separation is to utilize a detector which is itself capable of automatically separating figure from ground. Such an approach motivated the original architecture of Carpenter and Grossberg (1988). There it was assumed that the detector consists of pairs of laser radars whose outputs are and-gated to separate figure from ground. For example, a range detector focussed at the distance of the figure can extract the figure and a contiguous piece of the ground. The figure can be detached from the ground by spatially intersecting the range pattern with a pattern from another detector that is capable of differentiating figure from ground. A doppler image can be intersected with the range image when the figure is moving. The intensity of laser return can be intersected with the range image when the figure is stationary (Gschwendtner, Harney, and Hull, 1983; Harney, 1980, 1981; Harney and Hull, 1980; Hull and Marcus, 1980; Kolodzy, 1987; Sullivan, 1980, 1981; Sullivan, Harney, and Martin, 1979).

More generally, arrays of laser radar detectors may be used to separate image figures into distinct network levels, or slabs. This can be accomplished by intersecting the output signals from multiple detectors that simultaneously inspect the image. For example, a series of range detectors can register all objects at a regular series $D, 2D, 3D, \ldots, ND$ of distances, within some tolerance ΔD; a series of doppler detectors can register all objects at a regular series $S, 2S, 3S, \ldots, MS$ of speeds, within some tolerance ΔS; and an $N \times M$ matrix of intersection images can be generated which extract the figure at each combination of distance iD and speed jS within this tolerance. These $N \times M$ intersection images can be processed simultaneously in parallel with CORT-X and invariant filters before they activate a parallel array of ART or ARTMAP pattern recognition architectures.

3. Figure-Ground Separation in FACADE Theory

During biological vision, the retinal detectors do not, in themselves, separate figure from ground. One task of neural network research is to suggest how subsequent network processes which are activated by the retinal detectors may generate this competence. Grossberg (1987; reprinted in Grossberg, 1988) has, for example, introduced a neural theory of binocular vision in which image figures are separated from one another into distinct network levels, or slabs. A macrocircuit of this theory is shown in Figure 4, where the vertically hatched boxes form part of the Boundary Contour System (BCS) and the dotted boxes form part of the Feature Contour System (FCS). The theory describes how parallel and hierarchical interactions between the BCS and FCS generate a multiplexed, multiple-scale representation, called a FACADE representation, of the scene's Form-And-Color-And-DEpth. Within this representation, figural components which encode distinctive combinations of features, such as prescribed combinations of color, depth, and size, are segregated from one another into different network levels. These levels, in turn, activate subsequent stages of network processing that are designed for visual object recognition (Figure 5).

For present purposes, the main insight that may be derived from FACADE theory is that a properly designed sequence of FCS-BCS-FCS operations can separate figure from ground. Henceforth all networks that use this strategy, including the present model, will be called *FBF networks*. To indicate how such a sequence of operations can separate figure

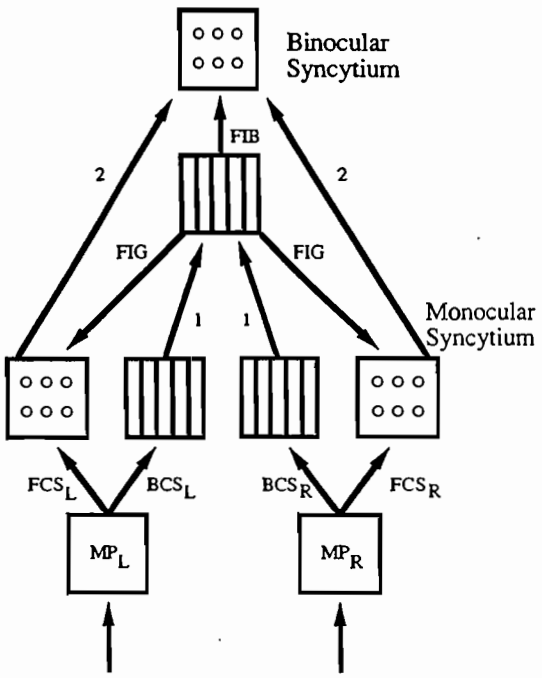

Figure 4. Macrocircuit of monocular and binocular interactions within the Boundary Contour System (BCS) and the Feature Contour System (FCS): Left and right monocular preprocessing stages (MP_L and MP_R) send parallel monocular inputs to the BCS (boxes with vertical lines) and the FCS (boxes with three pairs of circles). The monocular BCS_L and BCS_R interact via bottom-up pathways labelled 1 to generate a coherent binocular boundary segmentation. This segmentation generates output signals called filling-in generators (FIGs) and filling-in barriers (FIBs). The FIGs input to the monocular filling-in domains, or syncytia, of the FCS. The FIBs input to the binocular filling-in domains, or syncytia, of the FCS. Inputs from the MP stages interact with FIGs at the monocular syncytia where they select those monocular FC signals that are binocularly consistent. The selected FC signals are carried by the pathways labelled 2 to the binocular syncytia, where they interact with FIB signals from the BCS to generate a multiple scale representation of form-and-color-and-depth within the binocular syncytia.

from ground, we review two different competences of FACADE theory: discounting variable illumination and multidimensional fusion.

4. Discounting Variable Illumination and Filling-In

The theory provides an explanation of how variable illumination conditions are automatically discounted and used to trigger a filling-in process that completes a surface representation over image regions which are suppressed by the discounting process. A monocular version of this process was modelled by Cohen and Grossberg (1984) and Grossberg and Todorović (1988) to explain data about monocular brightness perception. This monocular

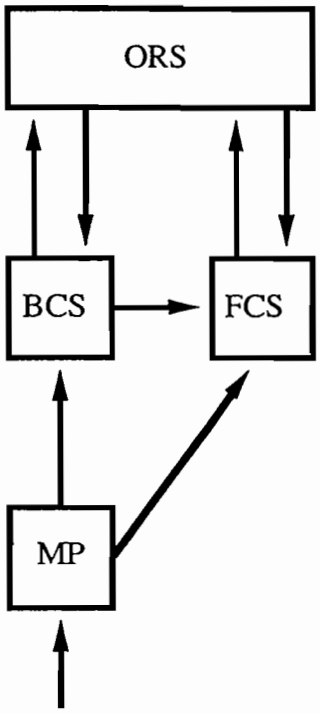

Figure 5. A macrocircuit of processing stages: Monocular preprocessed signals (MP) are sent independently to both the Boundary Contour System (BCS) and the Feature Contour System (FCS). The BCS preattentively generates coherent boundary segmentations from these MP signals. These structures send outputs to both the FCS and the Object Recognition System (ORS). The ORS, in turn, rapidly sends top-down learned expectation signals to the BCS. These expectations can modify the preattentively completed boundary structures using learned, attentive information. The BCS passes these modifications along to the FCS. The signals from the BCS organize the FCS into perceptual regions wherein filling-in of visible brightnesses and colors can occur. This filling-in process is activated by signals from the MP stage. The completed FCS representation, in turn, also interacts with the ORS. Fusion of BCS and FCS representations occurs in the ORS.

model is schematized in Figure 6.

In this model, variable illumination conditions are discounted by a shunting on-center off-surround network (Level 2), which constitutes the first FCS stage. (See the Appendix for all equations.) Image regions of high relative contrast are amplified and regions of low relative contrast are attenuated as a consequence of the discounting process. The shunting network, in turn, topographically activates a filling-in network (Level 6) which constitutes the second FCS state. This filling-in network uses a nonlinear diffusion process to complete a brightness representation over both the amplified and attenuated image regions.

Filling-in is restricted to compartments whose boundaries are defined by topographic signals from the Boundary Contour System, or BCS (Levels 2–5). The BCS converts signals

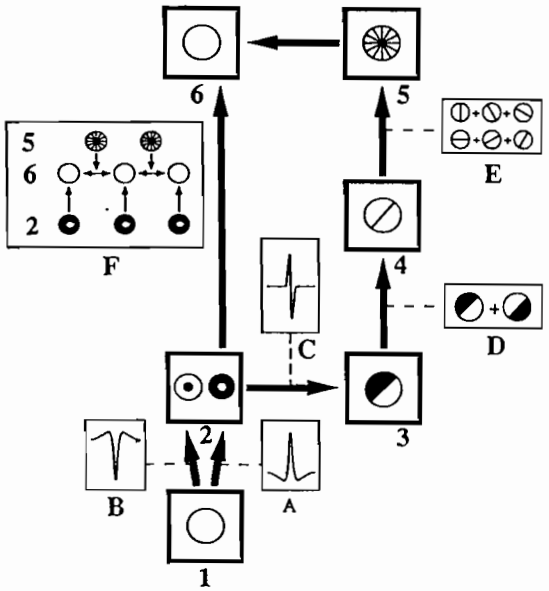

Figure 6. Grossberg-Todorović monocular model of how the FCS discounts variable illuminants and regulates featural filling-in: The thick-bordered rectangles numbered from 1 to 6 correspond to the levels of the system. The symbols inside the rectangles are graphical mnemonics for the types of computational units residing at the corresponding model level. The arrows depict the interconnections between the levels. The thin-bordered rectangles coded by letters A through E represent the type of processing between pairs of levels. Inset F illustrates how the activity at Level 6 is modulated by outputs from Level 2 and Level 5. This simplified model directly extracts boundaries from image contrasts, rather than generating emergent segmentations from image contrasts. The model's key elements concern how the Level 2 network of shunting on-center off-surround interactions discounts variable illuminants while extracting Feature Contour signals, and how Level 5 fills-in these signals via a nonlinear diffusion process within the compartments defined by BCS output signals.

from the first FCS stage (Level 2) into a boundary segmentation one of whose functions is to trigger a BCS-FCS interaction that contains the filling-in process at the second FCS stage. The result of this FBF interaction is a surface representation of featural quality, such as brightness or color, that is relatively uncontaminated by illumination conditions.

Figures 7 and 8 summarize computer simulations of Grossberg and Todorović (1988) that illustrate how the illuminant is discounted in FCS Level 2, and how the subsequent BCS-FCS interaction at Level 6 controls the filling-in process that completes the brightness representation. The image schematized in Figure 7a is called a McCann-Mondrian (Land, 1977). It is a patchwork of rectangular regions each with a different luminance. The image is uniformly illuminated. In Figure 7a, each circle's radius is proportional to the luminance registered by a network node located at the center of the circle. Figure 7b represents the activation pattern of the shunting on-center off-surround network at Level 2, Figure 7c represents the boundary representation at Level 5. Only on-cells were used to generate this

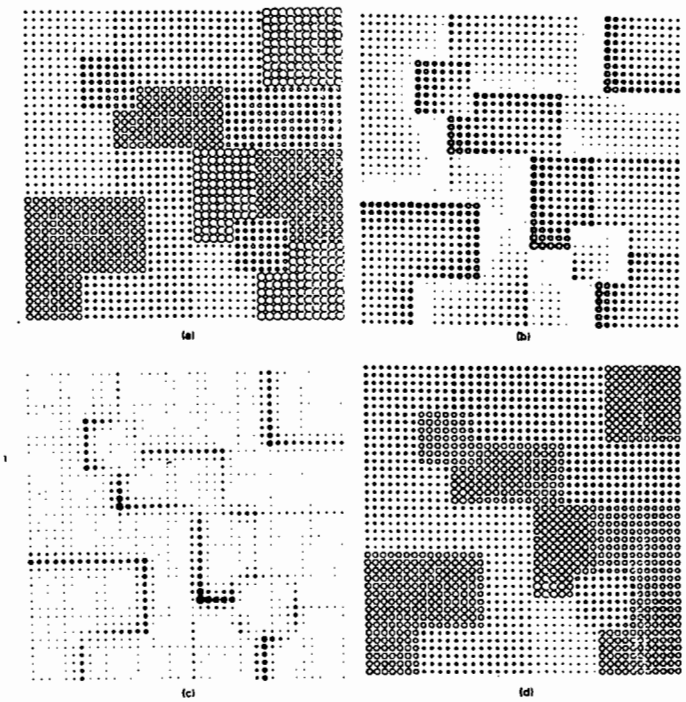

Figure 7. The evenly illuminated Mondrian. (a) The stimulus distribution consists of 13 homogeneous polygons with 4 luminance levels. Note that the square in the upper left portion of the stimulus has the same luminance as the square in the lower right portion. However, the average luminance of the regions surrounding the lower square is higher than the corresponding average luminance for the upper square. (b) The on-cell distribution. The amount of on-cell activity within the upper square is higher than within the lower square. (c) The Boundary Contour output. (d) The filled-in syncytium. The upper square is correctly predicted to look brighter than the lower square.

boundary, which consequently has uneven strength at concave and convex boundary shapes. Figure 7d represents the filled-in representation at Level 6. The diffusion spatially averages the activation patterns of Figure 7b within the compartments defined in Figure 7c.

In Figure 8a, the same image depicted in Figure 7a is illuminated from the lower right corner. Because the shunting on-center off-surround network at Level 2 effectively discounts the illuminant, the Level 2 activation patterns in Figures 8b and 7b are essentially identical. Hence the subsequent boundary patterns (Figures 8c and 7c) and filled-in patterns (Figures 8d and 7d) are also essentially identical.

In Figure 8d, the brightness, or activation level, of the square region in the upper left corner is larger than that of the square region in the lower right corner. In Figure 8a, the luminance, or activation level, of the upper left corner is smaller than that of the square region in the lower right corner. This luminance-to-brightness reversal compensates for the larger intensities of illumination in the lower right region.

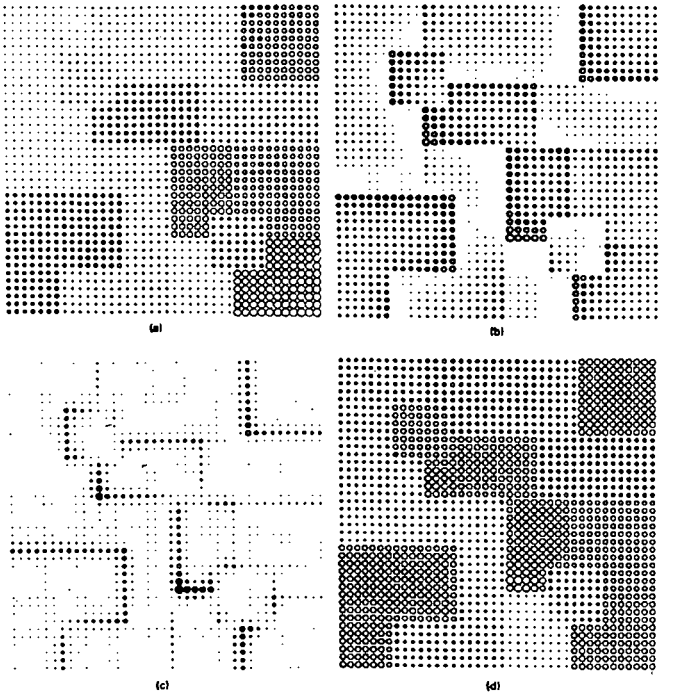

Figure 8. The unevenly illuminated Mondrian. (a) The stimulus distribution simulates the transformation of Figure 7a caused by the presence of a light source whose intensity decreases linearly from the lower right corner toward the upper left corner of the stimulus. The lower square is now more luminant than the upper square. (b) The on-cell distribution. (c) The Boundary Contour output. (d) The filled-in-syncytium. Figures 8b, 8c, and 8d are very similar to the corresponding figures for the evenly illuminated Mondrian (Figure 7). This illustrates the model's discounting of the illuminant. In addition, the upper square is still predicted to appear brighter than the lower square.

5. Multidimensional Fusion

An FBF interaction may be used to represent scenic form, notably scenic surface properties, as it separates figure from ground. This is achieved by embedding an FBF interaction into a binocular version of the theory. In the binocular theory, form, color, and depth are multiplexed together in the final representation; hence the mnemonic Form-And-Color-And-DEpth, or FACADE, in the theory's name.

The binocular version of the theory suggests how monocular image data from both eyes can be selectively processed so that only the binocularly consistent monocular data from each eye is allowed to influence the FACADE representation. Figure 4 schematizes the network that is used. In it, monocular BCS signals are derived from the monocular FCS patterns that discount the illuminant for each eye. These monocular BCS signals interact topographically to form the binocular boundary segmentation along the pathways labelled 1. This boundary segmentation regularizes and completes all the boundary data, across multiple spatial scales,

that are capable of being binocularly fused. Binocularly discordant, or rivalrous, data are suppressed by competitive interactions. The binocular boundary segmentation sends topographic signals, called *filling-in barriers* (FIBS), to the monocular filling-in networks within the FCS that correspond to the left eye (FCS_L) and the right eye (FCS_R). This BCS-FCS interaction allows only those monocular featural data from (FCS_L) and (FCS_R) that are consistent with the binocular boundary segmentation to generate topographic output signals, labelled 2, to the binocular FCS stage. This BCS-FCS interaction carries out a type of figure-ground separation, since an FCS region can generate output signals only if it is surrounded by binocular FIGS from the BCS.

The binocular FCS stage is called the binocular syncytium. In the binocular syncytium, the selected monocular FCS signals from both eyes interact once again with the binocular BCS signals. Here the surviving monocular FCS signals activate a filling-in process within the compartments that are defined by the binocular BCS signals. These BCS signals are called *filling-in barriers* (FIBS) because they contain the filling-in process within their boundaries. Both FIGs and FIBs obey the same equations. Their different effects are due to their action at different locations in the network hierarchy. The FACADE representation that is generated within the binocular syncytium completes the preattentive figure-ground separation process by grouping distinctive combinations of features into figures within separate network levels, or slabs. These slabs thereupon send adaptively filtered signals to subsequent processing levels for purposes of visual object recognition.

Within such a biological theory of vision, the process of separating figures into different slabs exploits the fact that the retina contains photodetectors with different spectral sensitivities; for example, three types of retinal cones and one type of retinal rod. The theory suggests how figures may be spatially parsed into separate slabs based, in part, upon the distinctive colors that are derived from these detectors. In addition, there exist multiple spatial scales and multiple binocular disparity computations within the theory that further parse figural components into separate slabs based upon different size-disparity correlations (Grossberg, 1987; Grossberg and Marshall, 1989). Thus, although FACADE theory uses an FBF network—actually an FBFBF network—to achieve figure-ground separation, this network exploits the existence of multiple detectors and multiple-scale reactions to these detectors to carry out the separation. For moving images, relative motion provides another measure that is used for figure-ground separation.

The present FBF model shows how figural components may be separated into separate slabs even if only a single detector is used that, in itself, cannot separate figure from ground, and if subsequent processing stages cannot use multiple scales, binocular interactions, or relative motion as cues for separation. For example, how can individual figures be separated from the cluttered ground of a picture taken with a camera that uses monochromatic film? We now show how a suitably designed FBF network can accomplish this task for at least certain classes of images. Section 6 provides an intuitive description of network stages and their effects. The Appendix describes network equations and parameters. These sections can be read in either order.

6. Figure-Ground Separation by a Monochromatic FBF Network: The Dye-Injected FBF

In the FBF model, a key assumption is that the filling-in process is activated by *internally*

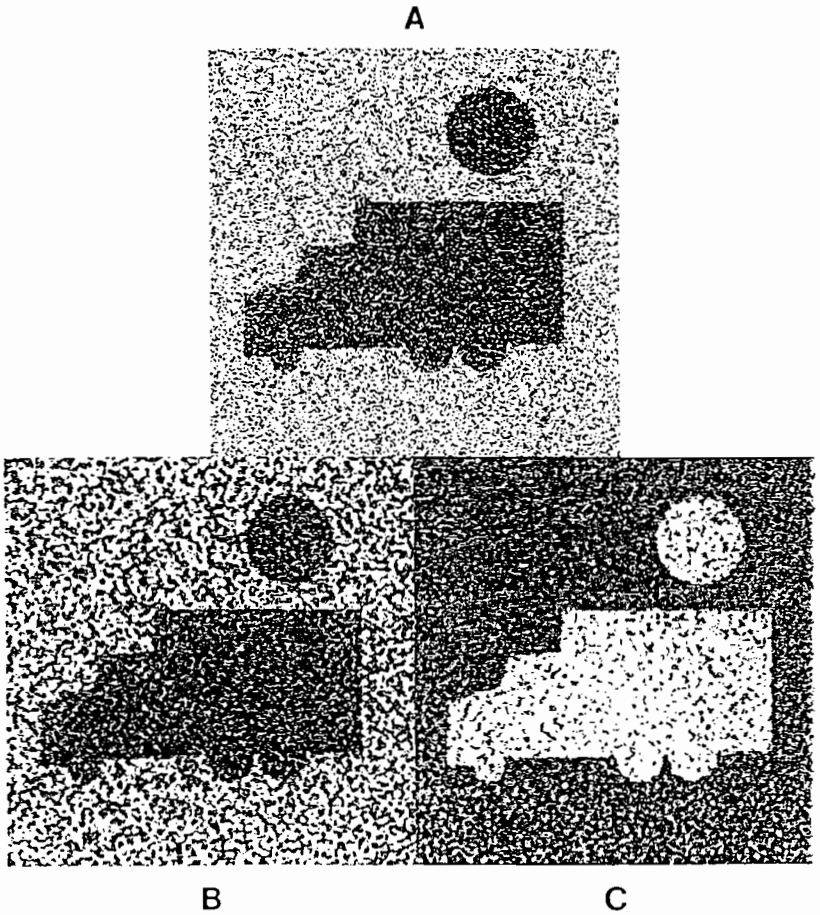

Figure 9. (a) The original figure in 50% noise (half the pixels are random). (b) The result of the ON-C shunting filter. (c) The result of the OFF-C shunting filter.

generated input sources. In particular, the network "paints" each connected figure of the image by using an internally generated "dye" that triggers the filling-in of that figure. This heuristic is realized by using the following procedure.

Step 1 (Discount the Illuminant). At the first FCS stage, variable illumination conditions are discounted both by a shunting on-center/off-surround network ("ON-C") and an off-center/on-surround network ("OFF-C"), operating in parallel. The ON-C network has a zero baseline activity. Hence, a cell's activity decays to zero if there is no signal within its entire receptive field. In contrast, the OFF-C network has a positive baseline activity. Because the OFF-C filter has a positive baseline activity and is inhibited by positive signal values, the network performs an image inversion. Both the ON-C network and the OFF-C network compensate for variable illumination and normalize the image using their shunting interactions (Figure 9).

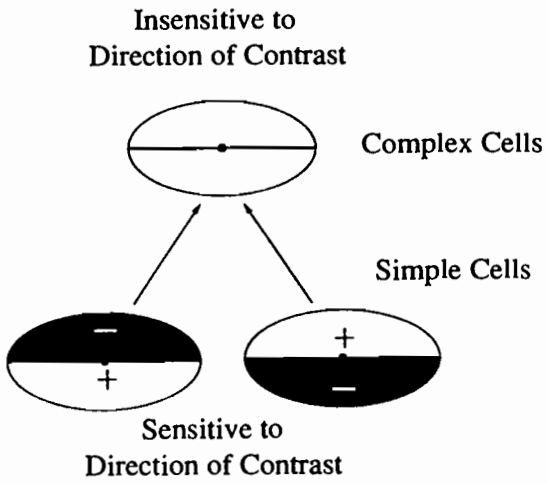

Figure 10. The simple cell and complex cell layers that process each of the two shunting network images in parallel. A horizontally oriented set of cells is shown. In Figures 11 and 12, only the outputs of horizontally oriented cells are displayed.

The ON-C and OFF-C networks operate in a complementary fashion. Along a straight boundary between a region of strong signal and one of no signal, both types of networks respond similarly by enhancing the contrast. At a concave corner of a high signal region, the ON-C network responds more strongly than the OFF-C network, while at a convex corner the converse is true (Grossberg and Todorović, 1988, Figure 6). This ON-OFF Complementarity Property also plays an important role in noise suppression when it interacts with the CORT-X 2 filter, as the next paragraphs explain.

Step 2 (CORT-X 2 Filter). The ON-C and OFF-C shunted images are transformed by a CORT-X 2 filter into a boundary representation. Each processing layer of this filter has the same number of cells as pixels in the image. The architecture is completely feedforward. Cells at a given layer have input fields ("IFs") that integrate over an area in the previous layer around its position in the field. Two separate scales (input field sizes) are used in parallel in the early stages of processing and are subsequently combined to take advantage of the best of their respective processing capabilities. The term *input field*, or in-field, is used instead of *receptive field* because the latter term from neurophysiology typically refers to the region at the first processing layer that influences the activity of a cell at any subsequent layer. Our layer-by-layer analysis of scale sizes requires a more microscopic analysis of network geometry.

The model's first stage, called the *simple cell layer*, is an oriented contrast detector that is sensitive to the orientation, amount, direction, and spatial scale of image contrast at a given image location. The orientation sensitivity is the result of an elliptically shaped IF (Figure 10). In-fields placed at equally spaced orientations operate in parallel at each position. The sensitivity to amount and direction of contrast is produced by exciting the cell with the inputs to one half of its IF, inhibiting it with the inputs to the other half, and thresholding the result to derive an output signal. The result is a half-wave rectification of

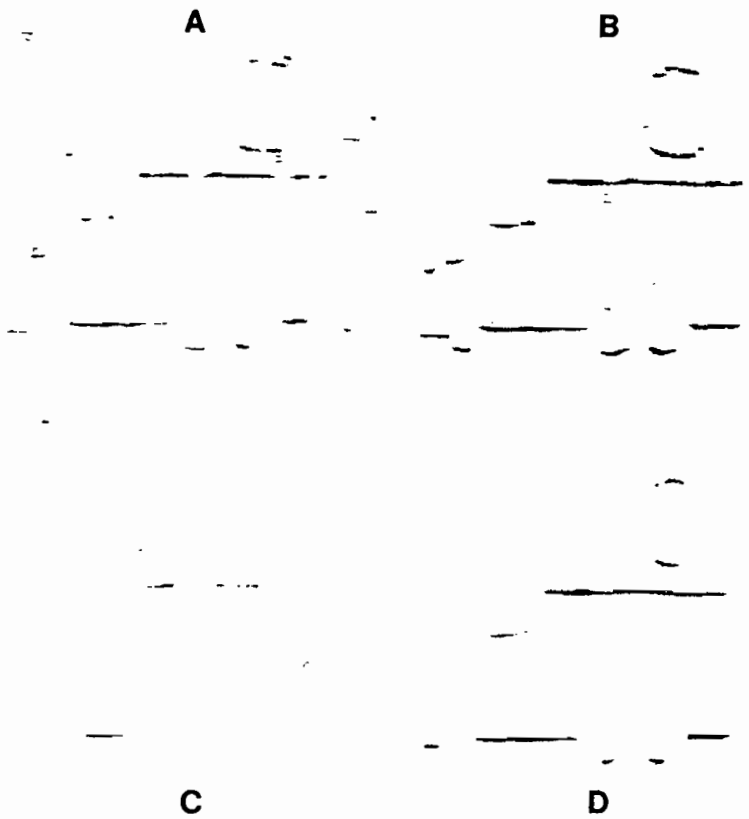

Figure 11. (a) Small scale horizontal complex cell output from the ON-C image of Figure 9b. (b) Small scale horizontal complex cell output from the OFF-C image of Figure 9c. (c) Large scale horizontal complex cell output from the ON-C image of Figure 9b. (d) Large scale horizontal complex cell output from the OFF-C image of Figure 9c.

the input.

The ON-C shunting network and the OFF-C shunting network each input to separate networks of simple cells at each receptive field size. Thus the ON-C and OFF-C networks together activate four networks of simple cells.

The outputs of these parallel simple cell networks are combined at each position to activate the second stage of the CORT-X 2 filter, called the *complex cell layer*. Complex cells are sensitive to the orientation, amount, and spatial scale of the contrast of the image at a given point, but not to the direction-of-contrast. This latter property is achieved by summing the outputs of all like-oriented simple cells at each position, including cells that are sensitive to opposite direction-of-contrast and that receive inputs from either the ON-C or OFF-C shunting networks. Two networks of complex cells, each sensitive to a different scale size, are generated in this way. Adding the half-wave rectified outputs from pairs of simple cells that are sensitive to opposite direction-of-contrast has the same net effect as full-wave

Figure 12. (a) ON-OFF small scale horizontal complex cell response derived by adding Figures 11a and 11b. (b) ON-OFF large scale horizontal complex cell response derived by adding Figures 11c and 11d.

rectification.

The complementary processing properties of the ON-C and OFF-C cells help to suppress noise when their outputs are further processed by simple cells and complex cells. To a first approximation, such a detector responds to the ratio of the inputs to each half of its oriented receptive field. A small amount of noise signal against a background of no signal affect them more than the same amount of "drop out" noise against a background of strong signal. The inversion of the image performed by the OFF-C filter changes the direction-of-contrast between any noise and its signal background. Thus, the noise will be disruptive in only one of the two parallel networks of simple cells and complex cells, while actual region boundaries will be strongly detected in both (Figure 11).

The output from both networks of contrast detectors are then summed, an operation which takes advantage of the ON-OFF Complementarity Property. First it yields roughly equal responses to both concave and convex curvatures. Second, since noise in a given region is suppressed in either the ON-C or the OFF-C network, while boundaries one are strong in both, the summation strengthens the boundary signals relative to the noise (Figure 12).

Figure 12 demonstrates that the two scales also exhibit another type of complementary processing capabilities (Carpenter, Grossberg, and Mehanian, 1989). The smaller scale filter does a better job of boundary localization than the larger scale filter, especially at positions of high boundary curvature, whereas the larger filter does a better job of noise suppression and boundary completion (Figure 13). In particular, the large scale filter achieves good noise suppression far from the image boundaries, but not within the radius of large scale IF's near these boundaries. The small scale filter is relatively poor at noise suppression anywhere. Canny (1986) has suggested how a single spatial scale can trade off between these virtues,

	SMALL SCALE	LARGE SCALE	SMALL + LARGE SCALE
BOUNDARY LOCALIZATION	*YES*	*NOT AT HIGH CURVATURE BOUNDARIES*	***YES***
NOISE SUPPRESSION	*NO*	*YES*	***YES***
BOUNDARY COMPLETION	*NOT AT SEGMENTS MISSING DUE TO NOISE*	*YES*	***YES***

Figure 13. Complementary processing properties of large and small oriented receptive fields. Reprinted from Carpenter, Grossberg, and Mehanian (1989) with permission.

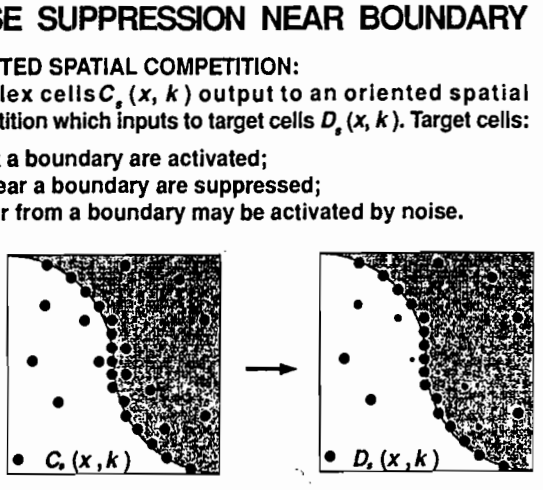

Figure 14. Oriented spatial competition inhibits noise pixels near the boundary. Reprinted from Carpenter, Grossberg, and Mehanian (1989) with permission.

but notes that "we cannot improve both simultaneously" (p. 684). The CORT-X family of models suggests a strategy whereby two or more scales can be combined to realize the best features of each, using the following operations.

The next stage of filtering is designed to control noise near boundaries. It converts complex cells into hypercomplex cells. Each complex cell excites the hypercomplex cell at the next level that corresponds to its position and orientation, while inhibiting hypercomplex cells at nearby positions that are not colinear with its axis of symmetry (Figure 14). This

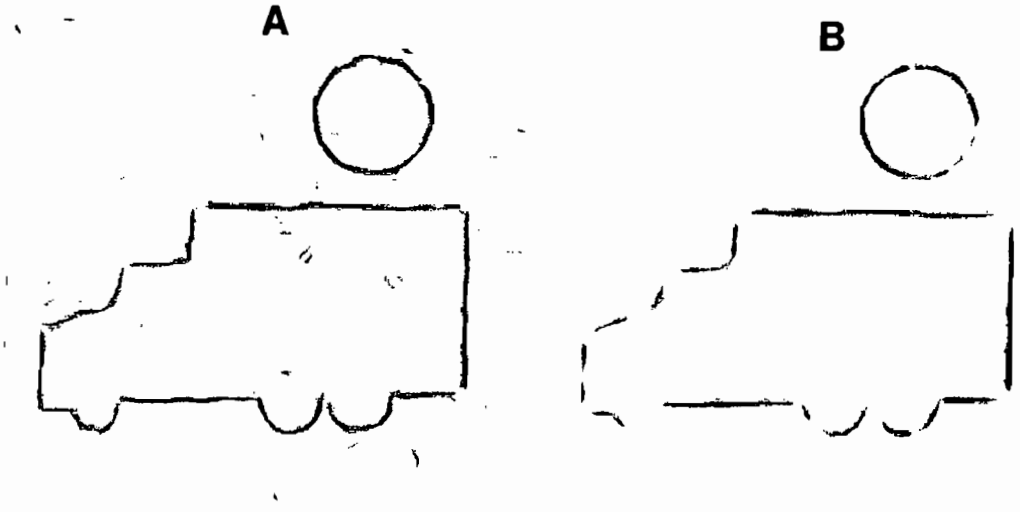

Figure 15. Output from the second competitive stage due to competition between orientations at each location. (a) Small scale. (b) Large scale.

interaction is called the first competitive stage. The level after that, is called the second competitive stage. It sharpens the activation pattern across orientations at each position and scale. In particular, at each position and scale, that cell is chosen whose orientation receives the maximal input from the first competitive stage (Figure 15).

The final operations include cooperative interactions between both filter sizes that select their desirable properties and eliminate their undesirable ones. The small scale's ability to localize boundaries and the large scale's ability to suppress noise and complete gaps in the boundaries are maintained by these cooperative interactions (Figure 16a).

Gaps in the boundary become more likely as the noise level of the original image increases. To overcome this problem, the cooperative level also includes oriented long-range interactions that are activated by the selected maximal responses in Figure 15. These long-range interactions activate an inactive cell if enough like-oriented cells at the previous level are active on both sides of the cell (Figure 16b). The cooperative cells play the role of the bipole cells in the Boundary Contour System (Grossberg and Mingolla, 1985a, 1985b). The final output of the CORT-X 2 filter is the sum of the combined-scales image and the completed-gap image (Figure 17).

Step 3 (Filling-In). The output from the CORT-X 2 filter is topographically mapped into M filling-in networks F_m, $i = 1, 2, \ldots, M$. In the Grossberg and Todorović (1988) article, the signals that trigger filling-in are generated by the image (Section 4). In the present application, they are generated by input sources that lie within the network. Moreover, each internally generated input is delivered to its filling-in network at a different position.

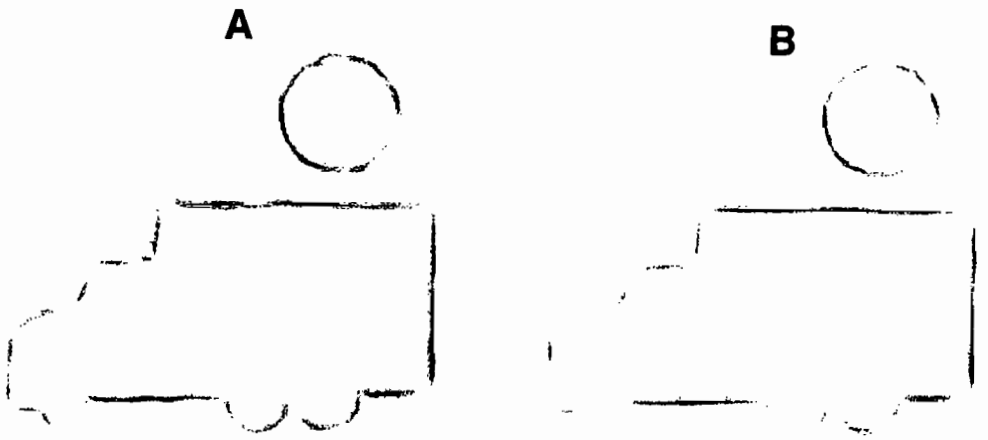

Figure 16. (a) Unoriented cooperation between both scales. (b) Oriented cooperation within the large scale.

Imagine, for definiteness, an $n \times n$ grid of $M = n^2$ nodes laid out over the boundary image generated by the CORT-X 2 filter. Each filling-in network F_m is associated with a different grid point where it will receive a featural "dye" injection into its copy of the boundary image. The injection then spreads unimpeded where there is no boundary signal, but does not spread through points where a boundary signal exists. Thus, each injection fills-in the connected figure that surrounds the injection point (Figure 18). Injection points do not need to occur at each pixel position. Rather, they should be dense enough to be enclosed, with high probability, by the smallest connected boundaries of regions that the CORT-X 2 filter can detect.

If the grid of injection points is dense enough, then all connected figural components will receive an injection within its boundary in at least one filling-in network F_m. This process is easy to replicate in large numbers because all the networks are identical except for the different, but regular, locations of the injected inputs, and all can operate independently and asychronously in parallel.

Step 4 (Figure-Ground Separation by a Double Opponent Network). Each filling-in network feeds its activation pattern in parallel to another pair of shunting networks, one on-center/off-surround (ON-C) and one off-center/on surround (OFF-C). This completes the second "F" operation in the FBF model. Because of their contrast enhancing and ratio-processing properties, these filters amplify the filled-in activity near figural boundaries while tending to suppress the low-contrast regions generated by the spreading of activation across the interiors of figures and background regions. In order to achieve figure-ground separation of a connected region whose boundary segmentation contains holes, the OFF-C output is subtracted from the ON-C output.

Such an operation generates cells that are called *double opponent* cells in neurobiology. The "double" opponency describes the two successive operations of inhibition, one in each

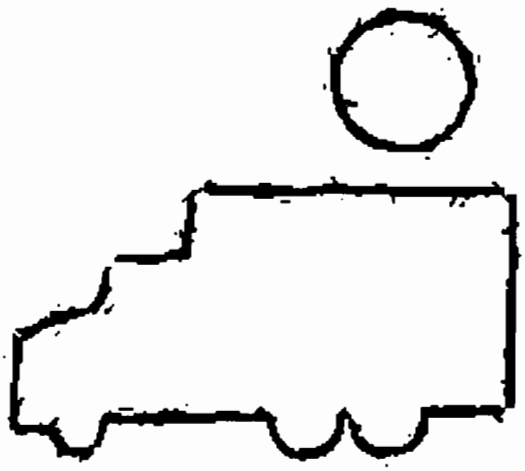

Figure 17. Final CORT-X 2 filter output.

shunting network and one between shunting networks, that defines the net output. Double opponent cells have traditionally been described as color processing cells *in vivo*. Grossberg (1987) noted that double opponency also helps to separate figures onto separate slabs in FACADE Theory. The FBF model adapts this observation to the single-detector single-scale case.

In order to understand how a double opponent network helps, consider first the case where there are no holes in a boundary at which the spreading activation could leak through. If the filling-in process has been given enough time to reach all of the enclosing boundary signals, then the ON-C filter produces an output only inside the enclosing boundaries, due to the injected activity (Figure 19E).

The OFF-C filter produces signal only outside the enclosing boundaries, due to the spontaneous baseline activity which has not been quenched by the input injection (Figure 19H). In this hole-free case, the ON-C filtered images effectively separate each connected region from all others in the original image. However, because of the Gaussian shapes of the kernels used in these filters, the edges in the ON-C patterns also exhibit a Gaussian spatial spread. Also, smooth gradients, such as those generated by activation spreading across regions where there are no boundary signals, may not be quenched to zero if the center and surround kernels are of unequal area. A better boundary can be generated by multiplicatively gating each ON-C pattern with the original CORT-X 2 output. The resulting M images, one on each slab, then contain among them all the separated figural boundaries of the original image (Figure 19J–L). This method can be run in real-time until the output from each copy generates a recognition event at its ART network, or the input image is removed.

If the original image contains so much noise that the CORT-X 2 filter is unable to produce boundaries without weak points or small gaps, then significant activity could leak out of figural components during the filling-in process (Figure 19A). On the other hand, the ART recognition process, operating in real-time, could recognize the figure before the

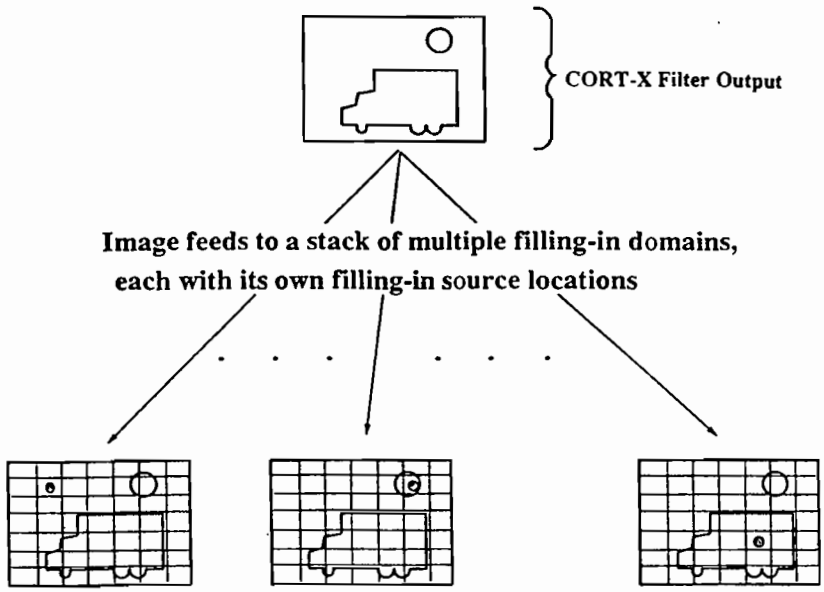

Figure 18. Copies of the CORT-X filter output are sent to M filling-in networks (three of which are pictured here). Activity is injected into a different place in each (gray disks) and begins to spread.

equilibrium state of equal activation on both sides of the boundary is reached.

Leakage of diffusing activity causes no problem if there are no other nearby object boundaries in the original scene. The spread of activation would produce a smooth gradient through the gap, and the shunting operation would not detect any contrast at the point of leakage or outside the object until well after the recognition event occurred (Figure 19J).

If, however, another object's boundary were near a boundary gap, then leaking activation followed by the ON-C shunting operation could detect the spurious boundary if the injection site were closer to the spurious boundary than to other boundaries of the figure.

In this case, the OFF-C filtered image is helpful (Figure 20). The ON-C signal at boundary regions exterior to the desired object is not as strong as the OFF-C filter signal at these points (Figure 20A), unless the boundary and the injection site are both proximal to the boundary gap. This property is due to the fact that very little injected activity would have spread there to excite the ON-C field, and the OFF-C field is tonically active. Subtracting the OFF-C image from the ON-C image to generate a double opponent image (Figure 20B), followed by CORT-X 2 gating (Figure 20C), therefore helps to suppress leakage from an incomplete boundary. The OFF-C subtraction does not distort the desired boundary signal when the ON-C signal is stronger than the OFF-C signal. In Figure 21, for example, a hole was made in the boundary of the truck near the boundary of another figure before filling-in occurs. The ON-C network clearly detects the outer boundary of the moon (Figure 21A). The OFF-C filtered signal is, however, stronger at these points (Figure 21B). Without taking the OFF-C network information into account, the output could cause difficulties for a

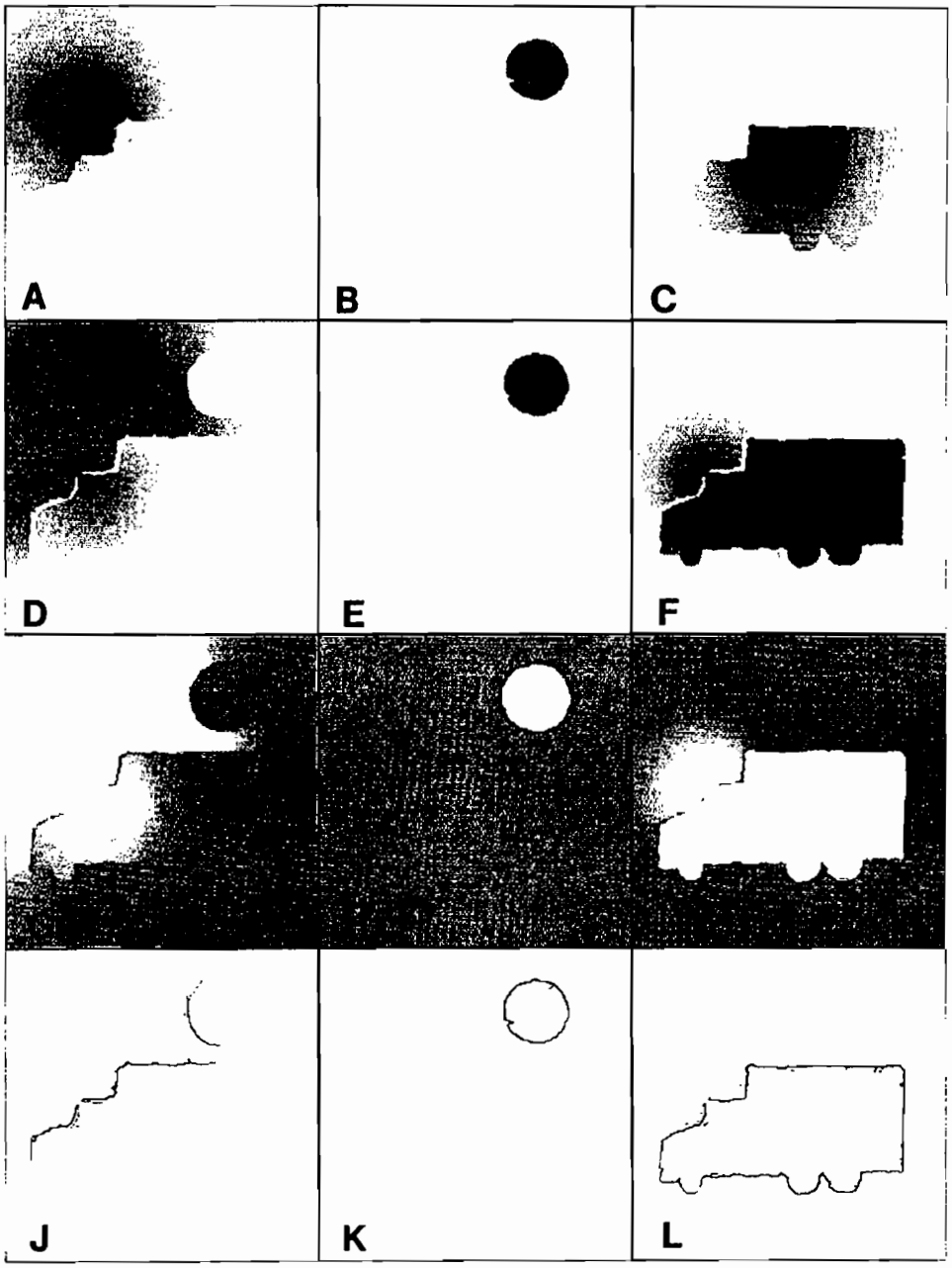

Figure 19. Row 1: Filling-in of activity is initiated at three different places in parallel filling-in networks. Due to the quantization of the gray scale, small filled-in activations do not print, even though they are detected by the shunting networks, as noted in rows 2 through 4. Row 2: The ON-C filter output of the respective filled-in regions. Row 3: The OFF-C filter output of the respective filled-in regions. Row 4: The final "separated figure" outputs to be passed along to a pattern recognizer such as an ART network.

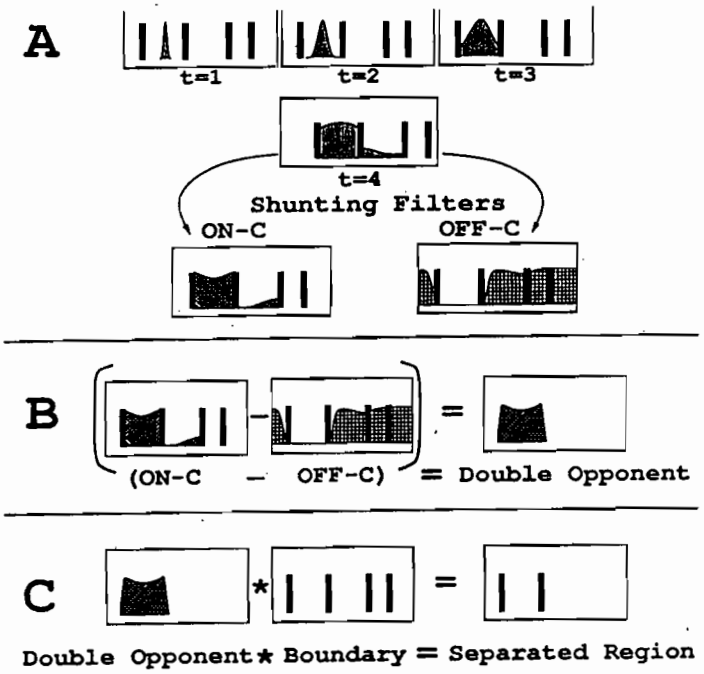

Figure 20. Each box represents a one-dimensional cross-section taken vertically through the truck/moon image of Figure 21. The 2 left-most black bars are from the truck boundary, the two rightmost, from the moon. (A) The injected activity at time=1 spreads, impeded only at locations where a boundary signal exists. At time t=4, we see leakage due to a nearby hole in the truck boundary which has flowed as far as the moon boundary. The ON-C and OFF-C filters of the activation contour are shown, with the boundary signals superimposed as black vertical bars. (B) The double opponency interaction suppresses the relatively weak leakage signal in the ON-C filter output. (C) The final gating operation with the original boundary from the CORT-X2 filter produces the desired representation of the separated truck figure.

pattern recognizer since the objects would not be separated (Figure 21C). Combining the two shunting network filters using a double opponent interaction produces the desired separation (Figure 21D).

A single figure is typically large enough to enclose several injection sites across the set of filling-in networks F_m. Thus, even if the injection site is close to a boundary gap and to a nearby spurious boundary in one network F_m, the injection sites of other filling-in networks will be further from the boundary gap. In some of these networks, double opponent processing can compensate for the gap, and trigger a correct recognition of the figure from the corresponding ART network. If the spurious boundaries are strong enough, as in Figure 20J, then the corresponding ART network will remain silent because the combination of partial figure and background is not similar enough to a previously learned recognition

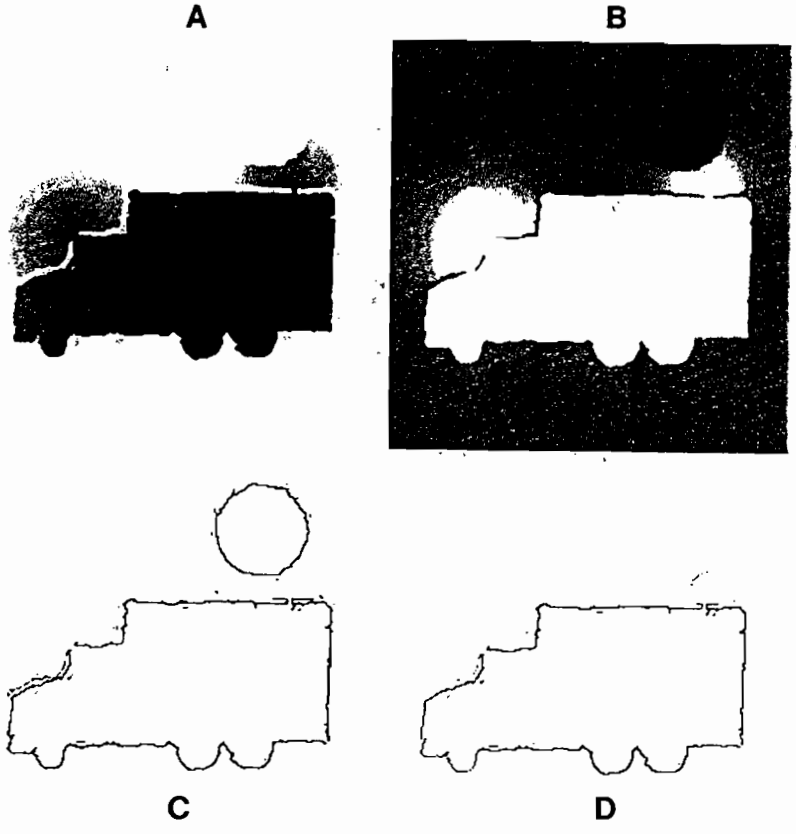

Figure 21. (a) The ON-C filter output of the filled in region with a hole in the boundary proximal to another figure. (b) The OFF-C filter of the filled in region. (c) The output as it would look without using the complementary OFF-C image. (d) The final image which uses the OFF-C filter to control the effects of the leakage leaving only the desired boundary signal.

code to be recognized. Parallel separation of multiple connected figures from other figures and background can thus be performed by an FBF architecture under noisy conditions by exploiting the parallel recognition by ART networks of boundary segmentations that are generated by sampling multiple filling-in perspectives.

In summary the FBF network separates figure-from-ground by using regular arrays of feedforward networks to discount the illuminant and to generate boundary segmentations, nearest-neighbor feedback signals for filling-in, and a proliferation of these circuits in parallel copies that input to parallel arrays of ART pattern recognition networks. The FBF networks thus seem to be appropriate designs for implementation in parallel hardware capable of operating at high rates in real time.

7. Recognition of Conjunctive Features

Why can the biologically-motivated FBF network automatically distinguish between the pair of connected and disconnected Minsky-Papert figures in Figure 3 that humans cannot distinguish? The main reason is that FBF networks use internally generated "dye injections". These inputs are topographically distributed across the entire perceptual space such that each dye injection is delivered to a different filling-in domain, or slab. In human perception, by contrast, the same Feature Contour signals that initiate Boundary Contour formation also act as input sources that trigger featural filling-in, as illustrated in Figures 6–8. The regions used by Minsky and Papert (1969) were all of the same color, of similar over-all shape, and occupied essentially the same region of their respective images. They would therefore tend to fill-in the same slab, or set of slabs, when they are being perceived by humans. Hence they could not be rapidly distinguished by filling-in a different set of slabs as a function of their connectivity.

This observation clarifies a recent controversy about human perception; namely whether target figures that differ from distractor figures by more than one type of feature can be separated from them by rapid parallel processing that does not require serial search. Treisman and her colleagues (Treisman and Gelade, 1980; Treisman and Souther, 1985) have suggested that such parallel processing can occur only if the target is distinguished from distractors along a single stimulus dimension, whereas if a target is defined by the conjunction of two or more stimulus dimensions, then it can only be separated from the distractors by a serial search process.

An exception to this rule was discovered by Nakayama and Silverman (1986), who showed that targets which differ from distractors by a combination of disparity and color, or of disparity and motion, can be rapidly separated without serial search. This result is consistent with the fact that FACADE representations of different disparity-and-color combinations activate different combinations of slabs (Grossberg, 1987). They are *structurally* separated in the representation, and hence can be rapidly detected.

In summary, the difficulty of distinguishing the connected and disconnected Minsky-Papert displays can now be explained by the same mechanisms that explain rapid search of Nakayama-Silverman displays, and that provide the heuristics for designing an FBF network for automatic figure-ground separation.

APPENDIX: FBF NETWORK EQUATIONS

Input Images

In the FBF model computer simulations, the images are 256 x 256 arrays with signal values in the interval [0,1]. The simulations pictured herein represent maximum signal strength by black and minimum signal strength by white. Noise was generated by randomly choosing a percentage of pixels and setting their values to a random number, or gray level, in the interval [0,1]. For the simulation pictured herein, 50% of the pixels were randomized. The input pattern $\{I_{ij}\}$ is thus represented as gray levels on a set of square pixels $\{P_{ij}\}$. Pixel P_{ij} attains the value I_{ij} at the set of image points $\{(u,v): i \le u < i+1, j \le v < j+1\}$.

Step 1 (Discount the Illuminant).

ON-C Network

Each node v_{ij} is placed, for notational convenience, at the center of the corresponding pixel P_{ij} where it receives input I_{ij}. The activity x_{ij} of the node v_{ij} at lattice position (i,j) obeys the shunting on-center off-surround equation:

$$\frac{d}{dt}x_{ij} = -Ax_{ij} + (B - x_{ij})C_{ij} - (x_{ij} + D)E_{ij} \tag{1}$$

where C_{ij} is the on-center interaction and E_{ij} is the off-surround interaction. Each C_{ij} and E_{ij} is a discrete convolution of the input pattern $\{I_{ij}\}$ with a Gaussian kernel (Figure 22A). Thus

$$C_{ij} = \sum_{p,q} I_{pq} C_{pqij} \tag{2}$$

and

$$E_{ij} = \sum_{p,q} I_{pq} E_{pqij} \tag{3}$$

where

$$C_{pqij} = C \exp\left\{-\alpha^{-2} \log 2[(p-i)^2 + (q-j)^2]\right\} \tag{4}$$

and

$$E_{pqij=} E \exp\left\{-\beta^{-2} \log 2[(p-i)^2 + (q-j)^2]\right\}. \tag{5}$$

In our simulations, $A = 134$, $B = 1$, $C = 7$, $D = .5$, $E = 3.333$, $\alpha = 1.3$, and $\beta = 1.875$. For this choice of parameters, the ON-C and OFF-C Gaussians are of equal area. At equilibrium $(dx_{ij}/dt = 0)$, (1) yields:

$$x_{ij} = \frac{\sum_{(p,q)}(BC_{pqij} - DE_{pqij})I_{pq}}{A + \sum_{(p,q)}(C_{pqij} + E_{pqij})I_{pq}} \tag{6}$$

OFF-C Network

The activity $\overline{x}_{ij}$ of the node v_{ij} at lattice position (i,j) obeys the shunting off-center on-surround equation:

$$\frac{d}{dt}\overline{x}_{ij} = -A(\overline{x}_{ij} - S) + (\overline{B} - \overline{x}_{ij})\overline{C}_{ij} - (\overline{=}x_{ij} + \overline{D})\overline{E}_{ij} \tag{7}$$

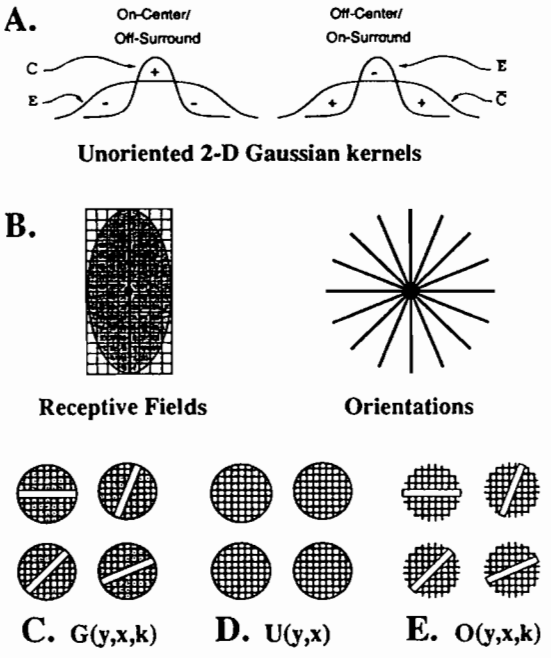

Figure 22. Kernels used in the CORT-X 2 filter. All oriented kernals have orientations at every $\pi/8$ radians.

where the on-center kernel of $\overline{x}_{ij}$ is the off-surround kernel of x_{ij}, and the off-surround kernel of $\overline{x}_{ij}$ is the on-center kernel of x_{ij}. In particular,

$$\overline{B} = D, \tag{8}$$

$$\overline{C}_{ij} = E_{ij}, \tag{9}$$

$$\overline{D} = B, \tag{10}$$

and

$$\overline{E}_{ij} = C_{ij}. \tag{11}$$

By Equations (8)–(11), (7) may be written as

$$\frac{d}{dt}\overline{x}_{ij} = -A(\overline{x}_{ij} - S) + (D - \overline{x}_{ij})E_{ij} - (\overline{x}_{ij} + B)C_{ij}. \tag{12}$$

At equilibrium,

$$\overline{x}_{ij} = \frac{AS + \sum_{(p,q)}(DE_{pqij} - BC_{pqij})I_{pq}}{A + \sum_{(p,q)}(C_{pqij} + E_{pqij})I_{pq}} \tag{13}$$

It follows by summing (6) and (13) that

$$x_{ij} + \overline{x}_{ij} = \frac{AS}{A + \sum_{(p,q)}(C_{pqij} + E_{pqij})I_{pq}} \tag{14}$$

which shows that for images $\{I_{pq}\}$ with constant Gaussianly filtered total activity

$$\sum_{p,q}(C_{pqij} + E_{pqij})I_{pq}, \tag{15}$$

the sum $x_{ij} + \overline{x}_{ij}$ is conserved and maintained at a positive value that increases with the tonic activity level S of $\overline{x}_{ij}$ in the dark. In our simulations, $S = .2$. The value for the parameter A (given the other parameter values) was chosen so that the activation of cell $\overline{x}_{ij}$ takes on values between S and 0 under spatially uniform illumination between 0 and 1 within its receptive field. Under such spatially uniform illumination conditions, $I_{pq} = I$ for all (p, q). Since for our choice of parameters the two Gaussian kernels are of equal area, let $\Phi = \sum_{(p,q)} E_{pqij} = \sum_{(p,q)} C_{pqij}$. Then Φ factors out so that the equilibrium equation (13) becomes

$$\overline{x}_{ij} = \frac{AS + I(D - B)\Phi}{A + 2I\Phi} \tag{16}$$

We derive the value for A by

$$A = \frac{(B - D)\Phi}{S} \tag{17}$$

so that when $I = 1$, the numerator vanishes. When $I = 0$, $\overline{x}_{ij} = S$ independent of A.

Step 2 (CORT-X 2 Filter).

All simple cell input fields are elliptical. Two sizes of input fields were used, indexed by the subscript s. The smaller scale, $s = 1$, had a major axis of 12 pixels and a minor axis of 6 pixels. The larger scale, $s = 2$, had a major axis of 20 pixels and a minor axis of 10 pixels. Orientations were chosen around the clock spaced by $\pi/8$ degrees. They are indexed below by the subscript k.

Simple Cells

A simple cell with index (i, j) is centered at the lower left hand corner of pixel P_{ij}. By this convention, the nodes v_{pq} of the shunting variables x_{pq} do not lie on the oriented axes that separate the excitatory and inhibitory halves of vertically and horizontally oriented receptive fields. The output of the pair of simple cells of scale s centered at index (i, j) with receptive field orientation k is defined by

$$S_{sL}(i, j, k) = \max[L_s(i, j, k) - \alpha_s R_s(i, j, k) - \beta_s, 0] \tag{18}$$

and

$$S_{sR}(i, j, k) = \max[R_s(i, j, k) - \alpha_s L_s(i, j, k) - \beta_s, 0] \tag{19}$$

where $L_s(i, j, k)$ and $R_s(i, j, k)$ are the contributions of the

left-half $l_s(i, j, k)$ and right-half $r_s(i, j, k)$, respectively, of the oriented input field; that is,

$$L_s(i, j, k) = \frac{\sum_{(p,q)\in l_s(i,j,k)} x_{pq} w_{pq}}{\sum_{(p,q)\in l_s(i,j,k)} w_{pq}} \tag{20}$$

and

$$R_s(i, j, k) = \frac{\sum_{(p,q)\in r_s(i,j,k)} x_{pq} w_{pq}}{\sum_{(p,q)\in r_s(i,j,k)} w_{pq}}, \tag{21}$$

where w_{pq} is a weighting factor equal to the proportion of the area of pixel P_{pq} (taken to be one square unit) covered by the receptive field. An activity x_{pq} was included in L_s or R_s if its pixel had a non-zero intersection with the corresponding half of the receptive field. Parameters α_s are threshold contrast parameters and parameters β_s are threshold noise parameters. We chose $\alpha_1 = 1.4$, $\alpha_2 = 2.0$, and $\beta_1 = \beta_2 = \beta = .012$. Each simple cell in (18) and (19) is sensitive to the opposite direction-of-contrast than its companion, as indicated by the indices L and R in S_{sL} and S_{sR}, respectively.

The ON-C and OFF-C networks each input to a different network of simple cells. The simple cells that receive signals from the ON-C network are denoted by S^+_{sL} and S^+_{sR}. The simple cells that receive signals from the OFF-C network are denoted by S^-_{sL} and S^-_{sR}.

The complex cells pool inputs from all simple cells of like orientation and scale that are centered at the same location, as described below.

Complex Cells

The complex cell output $C_s(i,j,k)$ is defined by

$$C_s(i,j,k) = F\left[S^+_{sL}(i,j,k) + S^+_{sR}(i,j,k) + S^-_{sL}(i,j,k) + S^-_{sR}(i,j,k)\right]. \tag{22}$$

Such a cell is sensitive to spatial scale s and amount-of-contrast centered at cell x with orientation k, but it is insensitive to direction-of-contrast. In our simulations, $F = .5$.

Hypercomplex Cells (First Competitive Stage)

The hypercomplex cells $D_s(i,j,k)$ at the first competitive stage

receive input from the spatial competition among the complex cells; that is,

$$D_s(i,j,k) = \max\left[\frac{C_s(i,j,k)}{\epsilon + \mu \sum_m \sum_{p,q} C_s(p,q,m) G_s(p,q,i,j,k)} - \tau, 0\right] \tag{23}$$

where $\epsilon = .1$, $\mu = 5$, $\tau = .01$. The oriented competition kernel $G_s(y,x,k)$ is normalized so that

$$\sum_{p,q} G_s(p,q,i,j,k) = 1. \tag{24}$$

As in Figure 22C, they are circular. Complex cells at the kernel periphery are weighted by the proportion of their area (taken to be one square unit) that are covered by the kernel. The grey areas in Figure 22C are inhibitory. Any cells whose defining pixel location lies within the one unit wide band through the middle of the kernel do not contribute to the inhibition. In our simulations, the small scale kernel is 8 units in diameter, and the large scale is 16 units in diameter.

Hypercomplex Cells (Second Competitive Stage)

The hypercomplex cells $D_s(i,j)$ at the second competitive stage realize a competition among the oriented activities $D_s(i,j,k)$ at each position x. For simplicity, the process is modelled as a winner-take-all competition; namely,

$$D_s(i,j) = D_2(i,j,K) = \max_k D_s(i,j,k), \tag{25}$$

where index K denotes the orientation of the maximally activated cell.

Multiple Scale Interaction: Boundary Localization and Noise Suppression

The interaction between scales is defined by the equation

$$B_{12}(i,j) = D_1(i,j) \sum_{p,q} D_2(p,q) U(p,q,i,j). \tag{26}$$

The unoriented excitatory kernel $U(p,q,i,j)$ is circular (Figure 22D), and normalized so that

$$\sum_{p,q} U(p,q,i,j) = 1 \tag{27}$$

and had a diameter of 8 units. All cells covered by the kernel contribute to the excitation to the extent that their area (taken to be one square unit) is covered by the kernel. The small scale hypercomplex cell $D_1(i,j)$ accurately localizes boundary segments and suppresses noise near the boundary. The large scale hypercomplex cell $D_2(p,q)$ supresses noise far from the boundary. The product $D_1(i,j)D_2(p,q)$ would simultaneously realize both constraints except that, due to the poor spatial localization of D_2, this term may be zero at boundary points of high curvature, thereby cancelling the good localization properties of $D_1(i,j)$. The effect of $D_2(p,q)$ in the equation is made more spatially diffuse by the kernel $U(i,j,p,q)$. The size of $U(i,j,p,q)$ is chosen to scale with that of $D_2(p,q)$ in order to compensate for the positional uncertainty of $D_2(p,q)$; a larger choice of $D_2(p,q)$ would necessitate a larger choice of $U(i,j,p,q)$. Although term $\sum_{p,q} D_2(p,q)U(p,q,i,j)$ in (26) localizes the boundary even less accurately than $D_2(p,q)$ does, the *product* of $D_1(i,j)$ with $\sum_{p,q} D_2(p,q)U(p,q,i,j)$ in (26) restores this loss of boundary localization. Moreover, kernel $U(p,q,i,j)$ causes no harm at locations p,q that are far from the boundary, since $D_2(p,q) = 0$ there (Carpenter, Grossberg and Mehanian, 1989).

Long-Range Cooperation: Boundary Completion

The function $B_{12}(i,j)$ represents the image boundary well except where boundary pixels are missing due to noise. More and larger boundary gaps are generated as the noise level increases.

The large detectors $D_2(i,j)$ can be used to partially overcome this problem. Because of the spatial uncertainty of the large detectors $D_2(i,j)$, they are capable of responding at locations where pixel signal strength has been reduced by noise. Such boundary signals may, however, be poorly localized. To overcome this tradeoff between boundary completion and localization, cooperative interactions among the large scale cells are defined by

$$B_2(i,j) = D_2(i,j) \max\Big[\sum_{p,q} D_2(p,q,K) O(p,q,i,j,K) - \delta, 0\Big]. \tag{28}$$

The oriented kernel $O(p,q,i,j,k)$ is defined by the one-unit-wide white strips in Figure 25E. Any cells with centers that lie within the one unit wide band contributes to the cooperative process. The kernel is normalized so that

$$\sum_{\text{(p,q) in kernel}} O(p,q,i,j,k) = 1. \tag{29}$$

In our simulations, the length of the kernel was 12 units, and $\delta = .001$.

CORT-X 2 Output

The final output of the CORT-X 2 filter is the rectified sum of the multiple scale interaction and the cooperative interaction:

$$B(i,j) = 1[B_{12}(i,j) + B_2(i,j)] \tag{30}$$

where

$$1[w] = \begin{cases} 1 & \text{if } w > 0 \\ 0 & \text{otherwise} \end{cases} \tag{31}$$

is the Heaviside function.

Step 3: Filling-In.

In each filling-in network F_m, an input is injected into a different area, with the shape of either a narrow Gaussian or a single node. For example, let $I = I(m)$ and $J = J(m)$ define the injection indices (I, J) of F_m. Then the injected input pattern to F_m was chosen to be

$$X_{ij}^{(m)} = X \exp\left\{-\gamma^{-2} \log 2\left[(I-i)^2 + (J-j)^2\right]\right\}. \tag{32}$$

This input pattern triggers filling in within F_m via the nonlinear diffusion equation (Grossberg and Todorović, 1988):

$$\frac{d}{dt} S_{ij}^{(m)} = -M S_{ij}^{(m)} + \sum_{p,q \in N_{ij}} (S_{pq}^{(m)} - S_{ij}^{(m)}) P_{pqij} + X_{ij}^{(m)}, \tag{33}$$

where $S_{ij}^{(m)}$ is the activity of the (i,j) node of F_m, and the index set N_{ij} of the summation contains the nearest neighbors of (i,j). The permeability coefficient P_{pqij} is defined by

$$P_{pqij} = \frac{\delta}{1 + \epsilon\Big(B(p,q) + B(i,j)\Big)} \tag{34}$$

where $B(p,q)$ and $B(i,j)$ are the outputs (30) from the CORT-X filter at the positions (p,q) and (i,j) respectively. Thus, activity spreads poorly, if at all, between cells where boundary signals are large, and easily where boundary signals do not exist.

In the simulations, the parameters $M = .0001$, $X = 50$, $\gamma = .5$, $\epsilon = 100000$, $\delta = 10$.

Step 4: Figure-Ground Separation.

Each filling-in network F_m inputs its filled-in image to shunting ON-C and OFF-C networks using the same equations (1) and (7) as above, with parameters $A = 1, B_{ij} = \overline{D}_{ij} = 1, C = 18, D_{ij} = \overline{B}_{ij} = .5, E = 3.333, \alpha = 2.96, \beta = 7$, and $S = .2$. Here the on-activations $x_{ij}^{(m)}$ and off-activations $\overline{x}_{ij}^{(m)}$ are parameterized by the filling-in network F_m from which they are derived. The boundary representation at position (i,j) of a figure derived from network F_m is defined by

$$R_{ij}^{(m)} = 1\left[(x_{ij}^{(m)} - \overline{x}_{ij}^{(m)}) B(i,j)\right] \tag{35}$$

where $1[w]$ is the Heaviside function. In other words, the figural boundary that is separated by network F_m is computed from the double opponent filter $(x_{ij}^{(m)} - \overline{x}_{ij}^{(m)})$ of filled-in F_m activation, gated by the CORT-X boundary segmentation $B(i,j)$.

Image Rendering

Each image has been scaled so that the maximum signals strength is mapped to black, and the minimum signal strength is mapped to white. Intermediate values map linearly onto a grey scale. The maximum signal strengths for the images are: Figure 9: (a) 1.0, (b) .125, (c) .197. Figure 11: (a) .041, (b) .067, (c) .018, (d) .058. Figure 12: (a) .054, (b) .035. Figure 14: (a) .261, (b) .219. Figure 15: (a) .013, (b) .038. Figure 16: (a) 1.0. Figure 18: Row 1 (col 1) 143.2, (col 2) 753.3, (col 3) 195.3; Row 2 (col 1) .324, (col 2) .330, (col 3) .325; Row 3 (col 1) .287, (col 2) .396, (col 3) .333; Row 4 (col 1) 1.0, (col 2) 1.0, (col 3) 1.0. Figure 19: (a) .325 (b) .332 (c) 1.0 (d) 1.0.

REFERENCES

Canny, J. (1986). A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **8**, 679–698.

Carpenter, G.A. and Grossberg, S. (1987a). A massively parallel architecture for a self-organizing neural pattern recognition machine. *Computer Vision, Graphics, and Image Processing*, **37**, 54–115.

Carpenter, G.A. and Grossberg, S. (1987b). ART 2: Stable self-organization of pattern recognition codes for analog input patterns. *Applied Optics*, **26**, 4919–4930.

Carpenter, G.A. and Grossberg, S. (1988). The ART of adaptive pattern recognition by a self-organizing neural network. *Computer*, **21**, 77–88.

Carpenter, G.A., Grossberg, S., and Mehanian, C. (1989). Invariant recognition of cluttered scenes by a self-organizing ART architecture: CORT-X boundary segmentation. *Neural Networks*, **2**, 169–181.

Carpenter, G.A., Grossberg, S., and Reynolds, J.H. (1991). ARTMAP: Supervised real-time learning and classification of nonstationary data by a self-organizing neural network. *Neural Networks*, **4**, 565-588.

Carpenter, G.A., Grossberg, S., Markuzon, N., Reynolds, J.H., and Rosen, D.B. (1991). Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps. Boston University Technical Report CAS/CNS-TR-91-016.

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

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

Grossberg, S. (1988). **Neural networks and natural intelligence**. Cambridge, MA: MIT Press.

Grossberg, S. (1990). Neural FACADES: Visual representations of static and moving Form-And-Color-And-DEpth. *Mind and Language*, **5**, 411-456.

Grossberg, S. and Marshall, J. (1989). Stereo boundary fusion by cortical complex cells: A system of maps, filters, and feedback networks for multiplexing distributed data. *Neural Networks*, **2**, 29–51.

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

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

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

Grossberg, S., Mingolla, E., and Todorović, D. (1989). A neural network architecture for preattentive vision. *IEEE Transactions on Biomedical Engineering*, **36**, 65–84.

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

Gschwendtner, A.B., Harney, R.C., and Hull, R.J. (1983). Coherent IR radar technology. In D.K. Killinger and A. Mooradian (Eds.), **Optical and laser remote sensing**. New York: Springer-Verlag.

Harney, R.C. (1980). Infrared airborne radar. **Proceedings of the IEEE 1980 electronic and aerospace systems conference**, 462–471.

Harney, R.C. (1981). Military applications of coherent infrared radar. In **Physics and Technology of Coherent Infrared Radar**. Proceedings of the SPIE, **300**, 2-11.

Harney, R.C. and Hull, R.J. (1980). Compact infrared radar technology. In **CO_2 laser devices and applications**. Proceedings SPIE, Vol. 227, 162–170.

Hull, R.J. and Marcus, S. (1980). A tactical 10.6 micrometer imaging radar. **Proceedings of the IEEE**. National Aerospace and Electronics Conference, Vol. 2, 662–668.

Kolodzy, P. (1987). Multidimensional machine vision using neural networks. In M. Caudill and C. Butler (Eds.), **Proceedings of the IEEE international conference on neural networks, II**, 747–758.

Land, E.H. (1977). The retinex theory of color vision. *Scientific American*, **237**, 108–128.

Minsky, M.L. and Papert, S.A. (1969). **Perceptrons**. Cambridge, MA: MIT Press.

Minsky, M.L. and Papert, S.A. (1988). **Perceptrons: Expanded edition**. Cambridge, MA: MIT Press.

Nakayama, K. and Silverman, G.H. (1986). Serial and parallel processing of visual feature conjunctions. *Nature*, **320**, 264–265.

Sullivan (1980). Infrared coherent radar. In **CO_2 laser devices and applications**. Proceedings SPIE, Vol. 227, 148–161.

Sullivan, L.J. (1981). Firepond laser radar. *Electro/81 Conference Record*, Session 34.

Sullivan, D.R., Harney, R.C., and Martin, J.B. (1979). Real-time quasi-3-diminsional display of infrared radar images. **Real-time signal processing II**. Proceedings of the SPIE, Vol. 180, 57–65.

Treisman, A. and Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology*, **12**, 97–136.

Treisman, A. and Souther, J. (1985). Search asymmetry: A diagnostic for preattentive processing of separable features. *Journal of Experimental Psychology: General*, **114**, 285–310.