



A Bayesian Framework for Understanding Texture Segmentation in the Primary Visual Cortex

TAI SING LEE*†

Received 22 December 1993; in revised form 21 December 1994

This paper presents a mathematical theory for understanding the computations involved in texture segmentation in the primary visual cortex. We propose that texture segmentation is a part of the early visual system's overall strategy to infer surfaces of objects in a visual scene. Based on this insight, we use the Bayesian inference paradigm to formulate the texture segmentation problem into a maximum *a posteriori* surface inference problem. The dynamical system for finding the optimal solution of this problem can be characterized by two concurrent and interactive processes: a gradual sharpening of the boundary signals and a simultaneous smoothing of the surface signals. The behavior of these dynamical processes was studied using both analytical and computational methods. We present some computational results and mathematical predictions. This theory suggests a novel framework for understanding the functional roles of the complex cells in the primary visual cortex.

Texture segmentation Surface interpolation Bayesian inference Neural model of V1

INTRODUCTION

Texture segmentation is a visual process that partitions a visual scene into regions of different textures. It is a preattentive and parallel process that covers a large area of the visual field simultaneously and is completed within 100 msec (Beck, Prazdny & Rosenfeld, 1983; Julesz, 1983; Treisman, 1985). The result of this global analysis is used to direct a small aperture of focal attention to interesting areas of the visual scene for further detailed analysis of form (Julesz, 1983).

In this paper, we present a Bayesian inference model for texture segmentation. This model is developed by Lee, Mumford and Yuille (1992). A model of similar spirit has also been developed by Geman, Geman, Graffigne and Dong (1990) using the Markov Random Field approach and directional residue statistical measures. In this paper, we analyze the dynamics of the model and discuss the relevance of this model to the understanding of the computation in the primary visual cortex.

Julesz (1983) suggested that preattentive texture segmentation arises from differential stimulation of different feature detectors in the brain. Based on this idea, Turner (1986), Fogel and Sagi (1989), Voorhees and Poggio (1988), Bovik, Clark and Geisler (1990), Malik and

Perona (1990) have developed successively more refined models for texture segmentation. All of these models involve several serial processing stages of filter response computation, Gaussian smoothing, and edge detection.

In our model, the surface interpolation process and the boundary detection process are combined into an interactive and concurrent system, with the resulting advantage of better boundary localization. Our model is the first model to use explicitly the idea of surface interpolation in texture segmentation. Furthermore, ours is also the first to explicitly assign probabilistic priors on texture variations in scale and orientation. This measure enables the model to tolerate deformation of texture within a surface due to perspective and surface shape of 3D objects.

This Bayesian approach yields a 'neural network model' which is similar in spirit to Grossberg and Mingola's (1985) Boundary Contour and Feature Contour systems proposed for brightness perception in the visual cortex. However, there are important differences between the two. One is that their feature system employs only passive diffusion, it is constrained by but not affecting the boundary system. In contrast, the boundary and surface systems in our model are tightly coupled and interactive. A much more important difference is that, because of the complexity of the computation, the simple gradient descent relaxation method their model utilizes would tend to trap the system in an inferior hypothesis. Our approach, on the other hand, utilizes a gradual relaxation strategy to seek better hypotheses.

*Division of Applied Sciences, Harvard University, Cambridge, MA 02138, U.S.A. [Email tai@hrl.harvard.edu.]

†Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139, U.S.A.

Our approach also differs from the other models in that we started with a computational theory that recognizes texture segmentation as a process for inferring surface region and surface boundaries, and then developed the algorithm and implementation to realize the theory. The basic biological insight provided by this model is that complex cells in the primary visual cortex should not be narrowly interpreted as local feature detectors but rather as computational units for interpreting surface and extracting surface boundaries. The goal of this paper is to propose a functional framework based on this model for understanding the complex cells in the primary visual cortex.

In the first part of the paper, we will present the Bayesian model for texture segmentation, deriving the basic properties of the dynamical system of the model and illustrating them with computational results. In the second part of the paper, we will discuss the biological and psychological evidence that lead us to suggest the possible existence of such a system in the brain.

BAYESIAN INFERENCE AND ENERGY FUNCTIONAL

The basic idea of Bayesian inference is that our interpretation of a visual scene is strongly influenced by our expectation and our prior knowledge of the world. The conditional probability of certain surface quality statistics f and location of surface boundary B given a measurement g is, by Bayes' rule,

$$P(f, B|g) = \frac{P(g|f, B)P(f, B)}{P(g)} \quad (1)$$

$P(g)$ is independent of f and B and is therefore a normalization factor. The maximum *a posteriori* estimate of $P(f, B|g)$ can be obtained by finding the f and B that maximize $P(g|f, B)P(f, B)$, where $P(g|f, B)$ is the constraint of f and B on data g , and $P(f, B)$ specifies the prior constraints on surface statistics f and surface boundaries B .

Julesz (1983) showed that some textures with identical first, second and third order global statistics can be preattentively discriminated. Therefore, the crucial determinants of texture discrimination are not global statistical measures but are the density of the so-called textons and their gradients.

Textons can be considered as descriptors of surface properties. Among the textons Julesz listed are oriented line segments, line terminators, crossings of line segments, and blobs. The simple cells in the primary visual cortex are ideal detectors of these textons (Hubel & Wiesel, 1968). Their receptive fields can be ideally modelled by the following self-similar family of Gabor wavelets (Lee, 1994),

$$\begin{aligned} \psi(x, y; \omega, \theta) = & \frac{2\omega}{5\sqrt{2\pi}} \exp\left(-\frac{\omega^2}{50}(2(x \cos \theta + y \sin \theta))^2\right) \\ & \cdot \exp\left(-\frac{\omega^2}{50}(-x \sin \theta + y \cos \theta)^2\right) \\ & \cdot [\exp(i(\omega x \cos \theta + \omega y \sin \theta)) - \exp(-\frac{25}{8})] \quad (2) \end{aligned}$$

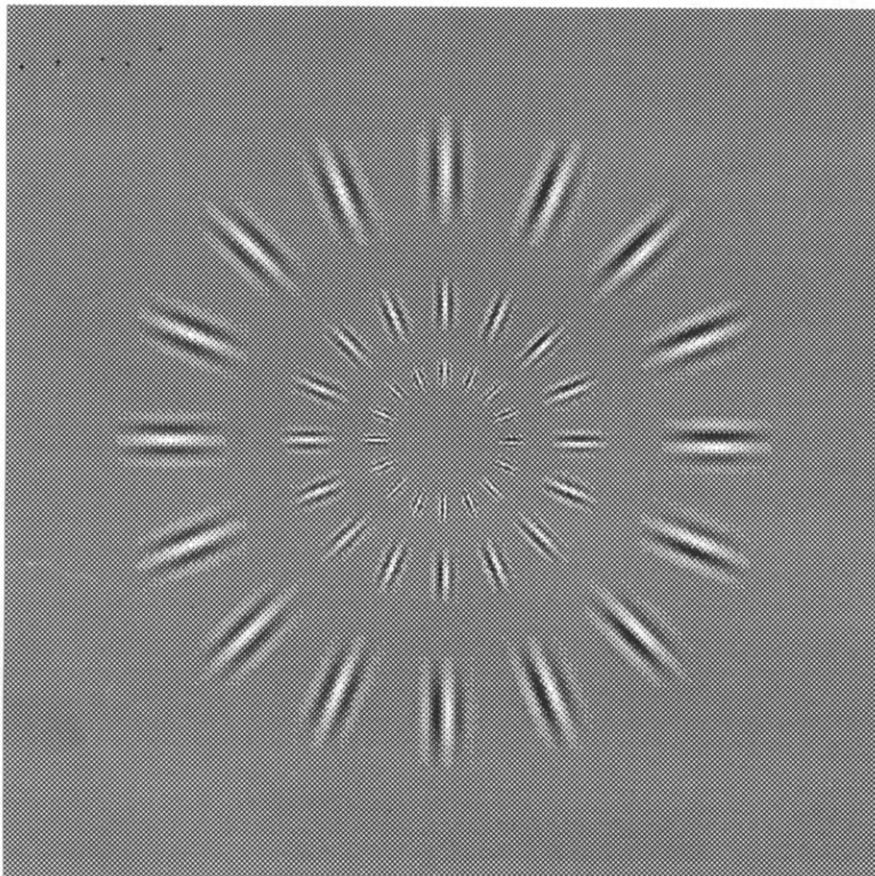
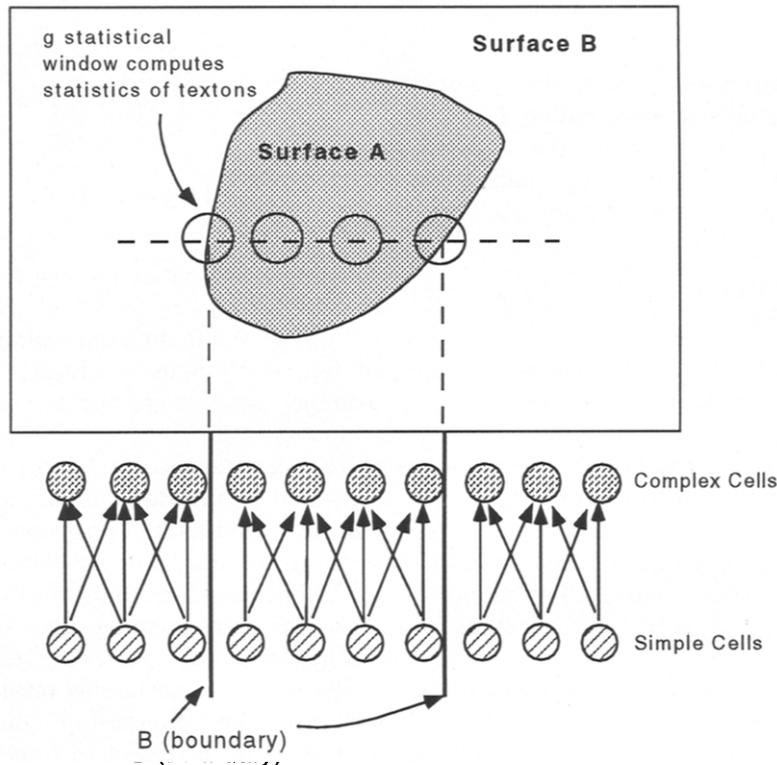


FIGURE 1. Receptive fields of an ensemble of simple cells can be modeled by a family of Gabor wavelets under biological constraints: 1.5 octave frequency bandwidth, 24 deg orientation bandwidth and 2:1 aspect ratio.

Direct Input to Complex Cells: Texton Statistics within B controlled window.



Boundary signals can cut away the lateral input of simple cells to the complex cells, i.e. modifying the g statistical window.

FIGURE 2. Complex cells compute the average response of the simple cells or the density of the textons within a g statistical window. The g statistical window is controlled by the boundary of the region. Four circular g windows are shown in the upper figure along the dashed line. The inner two windows compute the simple cells' responses within the full extent of the windows, while half of each of the outer two windows has been cut off by the boundary. In such cases, only the responses within the boundaries, i.e. the gray part of the windows, are used. The size of the g window can be controlled by the selective inactivation of the feedforward connection as shown in the lower figure.

where ω is the radial frequency and θ is the wavelet orientation. The Gabor wavelet is centered at $(x = 0, y = 0)$ and its L^2 norm is normalized to 1, i.e. $\langle \psi, \psi \rangle = 1$. The aspect ratio is 2:1, the frequency bandwidth is 1.5 octaves, and the orientation bandwidth is 24 deg. A discrete set of this family of wavelets is shown in Fig. 1.

The wavelet transform is given by, with

$$\sigma = \frac{2.5}{\omega},$$

$$\begin{aligned} \mathcal{W}I(x, y, \sigma, \theta) \\ = \iint I(x - x_0, y - y_0) \psi(x, y, \sigma, \theta) dx dy. \end{aligned} \quad (3)$$

$\mathcal{W}I$ at each x, y, σ, θ would be represented by the firing rates of four simple cells (positive and negative coefficients of the even and off symmetric wavelets) in the cortex.

Julesz's (1986) psychological data suggested that there is a Δ -neighborhood (at least twice the element size) in which texton densities and their gradients are computed.

Also, textons organize into texture percept only if their spatial separation is smaller than the diameter of a Δ -neighborhood window. This window scales with the Gabor wavelets and the texton density within it is given as follows:

$$\tilde{g}(x, y, \sigma, \theta) = \frac{1}{A} \iint_{\tilde{x}, \tilde{y} \in N_\sigma(x, y)} R(\mathcal{W}I(\tilde{x}, \tilde{y}, \sigma, \theta)) dA \quad (4)$$

where A is the area within a σ dependent neighborhood window N_σ centered at (x, y) . The spatial extent of the N_σ window within which the statistical average of $\mathcal{W}I$ is computed should be constrained by the boundary (see Fig. 2). However, this constraint is not considered in the present model. $R(\cdot)$ is either the rectification of (Malik & Perona, 1990) or the power modulus of the simple cell responses (Pollen, Gaska & Jacobson, 1989). Since Gabor wavelets are both feature detectors and frequency analyzers, $\tilde{g}(x, y, \sigma, \theta)$ is considered as the local distribution of texton statistics as well as of the spectral power in an image. The contrast normalization mechanism in the visual cortex (Heeger, 1992; Albrecht & Geisler,

1991) can be used to normalize this local distribution as follows,

$$g(x, y, \sigma, \theta) = \frac{\tilde{g}(x, y, \sigma, \theta)}{\iint \tilde{g}(x, y, \tilde{\sigma}, \tilde{\theta}) d\tilde{\sigma} d\tilde{\theta} + \eta} \quad (5)$$

where η is a constant.

The measured local distribution g , the relatively global estimate of the texture statistics distribution f of the surfaces and the estimate of the boundaries B of the surfaces are constrained by the following multivariate normal distribution, with standard deviation μ_d ,

$$P(g|f, B) = \frac{1}{Z} \prod_{x,y,\sigma,\theta} \exp\left(-\frac{1}{2\mu_d^2}(g-f)^2\right). \quad (6)$$

The priors (natural constraints) on the surface statistics f and boundaries B are summarized as follows:

- (1) Spatial homogeneity: statistics of surface qualities on the surface of an object tend to be continuous and similar.
- (2) Variation of texture in scale and orientation due to perspective and 3D surface shape will introduce a gradual and systematic shift in the statistical distribution.
- (3) An abrupt change in these statistics is indicative of occlusion boundaries between surfaces.
- (4) Any given visual scene is to be segmented into a limited number of surface regions with boundaries of a finite length.

Prior 1 is formulated probabilistically by assuming partials f_x and f_y within a surface follow two i.i.d. normal distributions. Prior 2 is formulated probabilistically by assuming partials $f_{\log\sigma}$ and f_θ within a surface also follow two i.i.d. normal distributions. Prior 3 is the prior that produces segmentation at locations of large texton gradient. Prior 4 is the prior specifying the amount of segmentation boundaries allowed within an image.

For homogeneous and stationary texture, the standard deviation of the global statistical distribution of g within a surface along each dimension is small. For very granular textures, the global statistic of g exhibits significant spatial fluctuation and its distribution will have a large variance. Similarly, for nonstationary texture such as pebbles on a beach in perspective or a wheat field swaying in the wind, global statistics of g within the surface will also have large variance in scale and orientation. Large variance in g along certain a dimension requires large smoothing force applied to f , i.e. assuming a small variance in the distribution of partial ∂f along the corresponding dimension.

The exponents of the prior constraints and the data constraints can be formulated into the following continuous energy functional,

$$E(f, B|g) \propto -\log(P(f, B|g)) \\ = \iiint \int_{(D \times \hat{D})} [f(\sigma, \theta, x, y)$$

$$-g(\sigma, \theta, x, y)]^2 d \log \sigma d\theta dx dy \\ + \iiint \int_{(D-B) \times \hat{D}} \left[\left(\gamma_\sigma \frac{\partial f}{\partial \log \sigma} \right)^2 \right. \\ \left. + \left(\gamma_\theta \frac{\partial f}{\partial \theta} \right)^2 + \left(\lambda_x \frac{\partial f}{\partial x} \right)^2 \left(\lambda_y \frac{\partial f}{\partial y} \right)^2 \right] \\ d \log \sigma d\theta dx dy + \alpha \int_B ds \quad (7)$$

where D is the spatial domain, \hat{D} is the spatial frequency domain; $(x, y) \in D$, $(\sigma, \theta) \in \hat{D}$. B is a set of C^1 curves cutting up D into different regions. f and g are functions of four continuous variables x, y, σ and θ . g is the average wavelet responses within a σ -dependent window.

The parameters $\lambda_x, \lambda_y, \gamma_\sigma, \gamma_\theta$, are inversely proportional to the standard deviations of g within each surface region along dimensions x, y, σ, θ , normalized with respect to μ_d . The smoothing effect they produce in each dimension introduces tolerance to signal variations along the corresponding dimension.

The maximum *a posterior* estimate of f and B , which is the *texture segmentation* result of the image, can be obtained by minimizing this energy functional $E(f, B|g)$, with respect to f and B .

The three integrals in equation (7) represent three competing forces in action: the force of reality (first term) demands fidelity to data; the force of homogeneity (second term) smooths out variation within a region and splits apart significantly different regions; and the force of tolerance (third term) prevents excessive splitting due to over-sensitivity, i.e. it controls the formation of breaks. These three forces act in concert to arrive at a balanced compromise according to the assumed prior statistical distributions specified by the parameters.

OPTIMIZATION DYNAMICS AND SYSTEM BEHAVIORS

The battle of these forces are best illustrated by a dynamical system that can be used to find the maximum *a posterior* estimate f and B . Since the energy functional is not convex, a simple gradient descent strategy will have the system trapped in a local minimum. A strategy to overcome this problem is called gradual relaxation: a convex functional is developed to approximate the real functional and it is slowly transformed back to the real functional in a series of successive stages of decreasing convexity. At each stage, the estimates of f and B are found by gradient descent. These estimates are used as starting points for the subsequent gradient descent in the next stage.

The energy functional [equation (7)] can be formulated as a series of approximating functionals E_κ , the deformation of which is controlled by a continuation

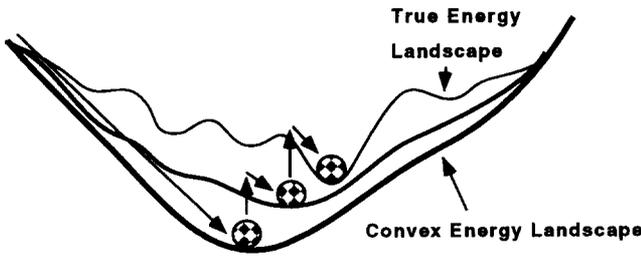


FIGURE 3. The energy landscapes of a sequence of deforming energy functionals and the successive gradient descent of a 'ball' in this sequence of landscapes: the system first converges to the global minimum of the convex landscape, which is then used as the starting point for the next descent in the new energy landscape. This strategy of successive gradual relaxation will allow the system to converge to a state that is close to a global minimum of the original energy functional (adopted from Blake & Zisserman, 1987).

parameter κ , in an approach similar to that taken by Ambrosio and Tortorelli (1990) and Richardson (1990).

$$E_\kappa = \iiint \iiint_R \left[(f - g)^2 + \gamma_\sigma^2 \left(\frac{\partial f}{\partial \log \sigma} \right)^2 + \gamma_\theta^2 \left(\frac{\partial f}{\partial \theta} \right)^2 + \lambda^2 \|\nabla f\|^2 (1 - l)^2 + \kappa \alpha \|\nabla l\|^2 + \frac{\alpha l^2}{4\kappa} \right] d \log \sigma d\theta dx dy \quad (8)$$

where f , g , γ_σ , γ_θ , and α are defined earlier, and $\lambda = \lambda_x = \lambda_y$ (assuming isotropic distribution of g over space). The earlier binary boundary B is replaced by a continuous value boundary variable $l(x, y)$. $\|\nabla f\| = \sqrt{(\partial f/\partial x)^2 + (\partial f/\partial y)^2}$, and similarly for $\|\nabla l\| = \sqrt{(\partial l/\partial x)^2 + (\partial l/\partial y)^2}$. κ , the continuation parameter, starts at 1 and decreases slowly over time. In the limit as $\kappa \rightarrow 0$, E_κ converges to E_0 , the original energy functional. Figure 3 illustrates the idea of successive gradient descent in a series of deforming energy landscapes.

The descent equations for f and l for minimizing this energy functional can be obtained from the associated

Euler-Lagrange equations by setting

$$\frac{df}{dt} = \frac{\partial E_\kappa}{\partial f} - \frac{\partial}{\partial x} \frac{\partial E_\kappa}{\partial f'_x} - \frac{\partial}{\partial y} \frac{\partial E_\kappa}{\partial f'_y} - \frac{\partial}{\partial \sigma} \frac{\partial E_\kappa}{\partial f'_\sigma} - \frac{\partial}{\partial \theta} \frac{\partial E_\kappa}{\partial f'_\theta} \quad (9)$$

$$\frac{dl}{dt} = \frac{\partial E_\kappa}{\partial l} - \frac{\partial}{\partial x} \frac{\partial E_\kappa}{\partial l'_x} - \frac{\partial}{\partial y} \frac{\partial E_\kappa}{\partial l'_y} \quad (10)$$

i.e.

$$\begin{aligned} \frac{df(\sigma, \theta, x, y, \kappa, t)}{dt} = & r_f \{ -f(\sigma, \theta, x, y, \kappa, t) \\ & + g(\sigma, \theta, x, y, \kappa, t) \\ & + \gamma_\sigma^2 \frac{\partial^2 f}{\partial^2 \log \sigma} + \gamma_\theta^2 \frac{\partial^2 f}{\partial^2 \theta} \\ & + \nabla \cdot [\lambda^2 \nabla f (1 - l(x, y, \kappa, t))^2] \} \quad (11) \end{aligned}$$

$$\begin{aligned} \frac{dl(x, y, \kappa, t)}{dt} = & r_l \left\{ \alpha \kappa \nabla^2 l(x, y, \kappa, t) \right. \\ & \left. + \lambda^2 (1 - l) \|\nabla f\|^2 - \frac{\alpha l}{4\kappa} \right\}. \quad (12) \end{aligned}$$

The parameters r_f and r_l are positive rate constants which control the rate of descent. At each κ , as κ changes slowly from 1 to 0, the system relaxes to an equilibrium, i.e. df/dt and dl/dt are driven to 0.

These equations describe the dynamics of the system. They suggest that the optimization is achieved by two concurrent and interactive processes of boundary-contraction and surface-expansion mediated by the three forces described. The flow of computation in this model is illustrated in Fig. 4. As time evolves, the surface statistical signals f diffuse in all four dimensions (σ, θ, x, y) , but their diffusion in space is modulated by the boundary signal $l(x, y)$. Therefore, the diffusion is nonlinear because it would be slowed down or blocked by the boundary signals. On the other hand, the boundary signals l undergo gradual contraction over space. They are activated by the gradient in the surface signals $\|\nabla f\|$ and they contract spatially over

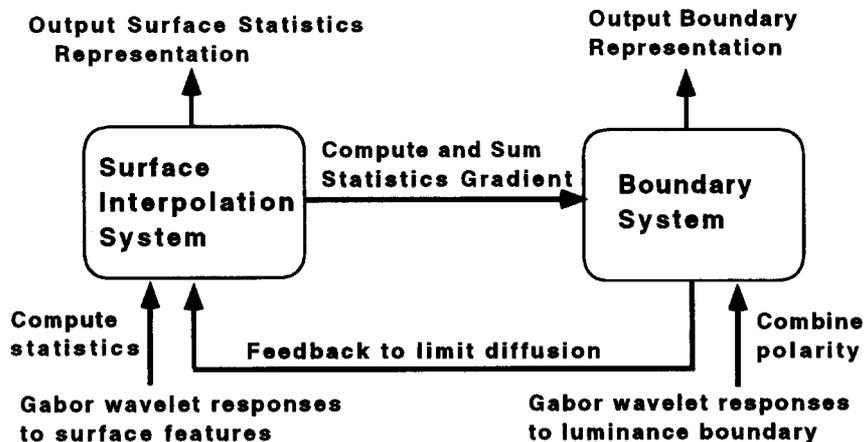


FIGURE 4. The computational flow of the model: the surface system computes the statistics of the simple cells and these statistical signals diffuse in both the spatial and spectral domains in the process of surface interpolation; the boundary system is activated by the gradient in the statistical signals and it feeds back to constrain the diffusion process in the surface system.

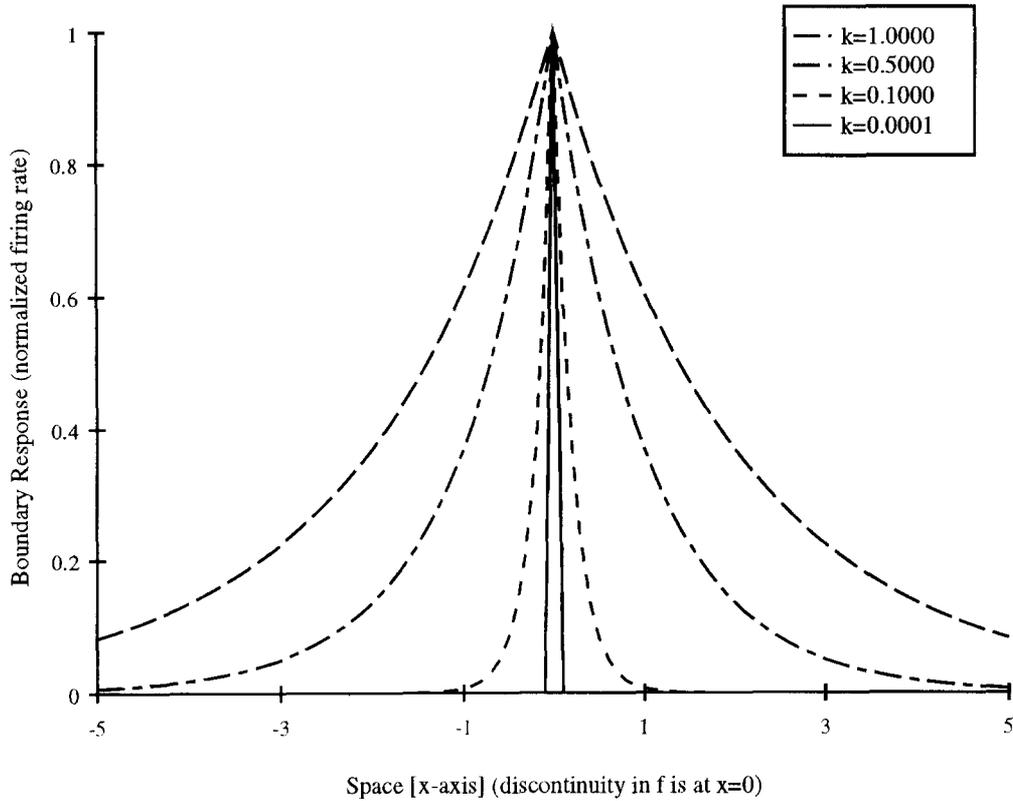


FIGURE 5. The spatial contraction of the boundary signals over time: in this image, $\nabla f \neq 0$ at $x = 0$ and $\nabla f = 0$ elsewhere. Initially, the boundary signal has a wide spatial spread. As κ decreases, the spatial spread of its response contracts toward the discontinuity at $x = 0$. The boundary responses at $\kappa = 1, 0.5, 0.1, 0.0001$ are shown in the figure.

time as $\kappa \rightarrow 0$. Therefore, the system is characterized by a *gradual sharpening* of the boundary signals as well as a simultaneous *nonlinear smoothing* of the surface signals.

The gradual contraction of the boundary signals can be understood by considering the system's response to a boundary. Let the system's responses in all texture statistical channels f be piecewise constant in x and y except at the boundary $x = 0$. In this case, the gradient of f along y is all zero, and the gradient along x is all zero except at $x = 0$; however, at $x = 0$, because of the collective gradient ∇f is nonzero and greater than the threshold, the boundary signal is active, i.e. $l(x = 0) = 1$. Therefore the second term in equation (12) just vanishes, and at the equilibrium of each κ relaxation stage (i.e. $dl/dt = 0$), the dynamical equation [equation (12)] yields

$$\alpha\kappa \frac{\partial^2 l}{\partial x^2} = \frac{l\alpha}{4\kappa} \Rightarrow l = e^{-\alpha|x|/2\kappa} \quad (13)$$

$\nabla^2 l$ controls the lateral interaction between the boundary signals l . As $\kappa \rightarrow 0$, the spatial extent of the lateral inhibition extends within the surface, and the boundary signal l contracts to a sharp line as shown in Fig. 5.

The smoothing of f is obvious from equation (11). The influence of the boundary process makes it a nonlinear diffusion equation, i.e. at the boundary, the diffusion is slowed down or interrupted. But far away from the boundary, the diffusion is linear. The dynamical equation becomes,

$$f - \lambda^2 \frac{\partial^2 f}{\partial x^2} - \lambda^2 \frac{\partial^2 f}{\partial y^2} - \gamma_\sigma^2 \frac{\partial^2 f}{\partial \log \sigma^2} - \gamma_\theta^2 \frac{\partial^2 f}{\partial \theta^2} = g(x, y, \log \sigma, \theta). \quad (14)$$

The predicted response of f at a particular point $(x_0, y_0, \sigma_0, \theta_0)$ in the spatial-spectral space, away from the boundary is given by convolving a Green's function with the input signal,

$$f(x_0, y_0, \log \sigma_0, \theta_0) = \iiint \iiint_{D \times \hat{D}} G(x - x_0, y - y_0, \log \sigma - \log \sigma_0, \theta - \theta_0) g(x, y, \log \sigma, \theta) + dx dy d \log \sigma d\theta \quad (15)$$

where C is a constant; $G(x, y, \sigma, \theta)$, the Green's function to equation (14), is given by,

$$G(x - x_0, y - y_0, \log \sigma - \log \sigma_0, \theta - \theta_0) = \frac{C}{r} K_1(r) \quad (16)$$

where K_1 is the modified Bessel function of the second kind of order 1 and

$$r = \sqrt{\left(\frac{\log(\sigma/\sigma_0)}{\gamma_\sigma}\right)^2 + \left(\frac{\theta - \theta_0}{\gamma_\theta}\right)^2 + \left(\frac{x - x_0}{\lambda_x}\right)^2 + \left(\frac{y - y_0}{\lambda_y}\right)^2} \quad (17)$$

is the distance away from the position $(x_0, y_0, \sigma_0, \theta_0)$ being considered.

For easier visualization, the Green's function in 1D is given by

$$G(x) = \frac{1}{2\lambda} e^{-|x|/\lambda}. \quad (18)$$

The implication of this analysis is that the process of surface statistics interpolation can be implemented in two ways: one is by the nonlinear diffusion of f , and the other is by the feedforward input from the simple cells with the weight (Green's function) of the feedforward connection dynamically controlled by the global statistical parameters $\lambda, \alpha, \gamma_\sigma, \gamma_\theta$ as well as by the boundary B .

COMPUTER EXPERIMENTAL RESULTS

The continuous energy functional [equation (7)] and its variation by the continuation method [equation (8)] allows us to analyze the dynamical system analytically and arrive at some basic understandings of the system with respect to boundary contraction and surface completion. We also implement the model in a parallel MASPAC computer to study its performance.

Both computer and biological implementation of the model requires discretization of the representation. Interestingly, the cortical sampling density in the spatial-spectral domain exceeds the minimum requirement (16 orientation columns, 4 frequency steps per octave, half receptive field overlap) for a tight frame which functions basically like a continuous representation (Lee, 1994). Therefore, the brain does have the representational machinery to implement the continuous energy functional. For computer implementation, because of the limit in computational resources, we have used a far coarser representation: 8 orientation steps and 3 scale steps with 1 octave apart. The frame is still relatively tight in this case, so that reasonable reconstruction can

be obtained using simple linear summation (see Fig. 6). Therefore, this coarse representation is a reasonable approximation of the continuous representation in our implementation.

A computational network has been set up to implement the discrete version of the continuous energy functional for computing the maximum *a posteriori* estimates of f and B . The architecture of computational implementation is shown in Fig. 7. The surface system is composed of a set of 24 lattices of nodes, each lattice corresponding to a wavelet channel of a particular orientation and frequency tuning. Each of these lattices (shown in Fig. 7 as the spatial lattice supported by the telephone poles) resembles Blake and Zisserman (1987)'s weak-membrane. The membranes in this model are coupled together to allow inter-membrane spectral diffusion so that the system can tolerate texture variation in scale and orientation. Therefore, this model has been called the *coupled-membrane* model (Lee *et al.*, 1992).

Many of the algorithms developed for minimizing the discrete version of the energy functional exploit the same strategy of gradual relaxation. Their dynamics (Blake & Zisserman, 1987; Geiger & Yuille, 1991) are qualitatively similar to the continuous dynamics discussed earlier. For discrete implementation in computer, we generalized the graduated nonconvexity algorithm of Blake and Zisserman to seek the best estimate of f and B . The computational results are used to evaluate the performance of the energy functional as well as to illustrate qualitatively the effect of boundary sharpening and surface spreading.

Figures 8 and 9 show the converged boundary signals of the texture system in response to a set of textured images. Responses to Mondrian and Brodatz demonstrated the system's ability to tolerate stochastic noise in granular textures. Responses to image Shell demonstrated the tolerance to gradual shifts in the orientation of the texture. Responses to Vase and Sierra demon-



Original Image

Reconstruction

FIGURE 6. Reconstruction of the standard image 'Lena': despite the coarse sampling density (8 orientations per cycle, 1 scale step per octave), the reconstruction resulting from simple linear summation of the product of the wavelets' coefficients and their receptive fields produce a nice interference-free reconstruction, suggesting that the representation is complete within a certain frequency band and is fairly continuous, thus can reasonably represent the input signals for our model.

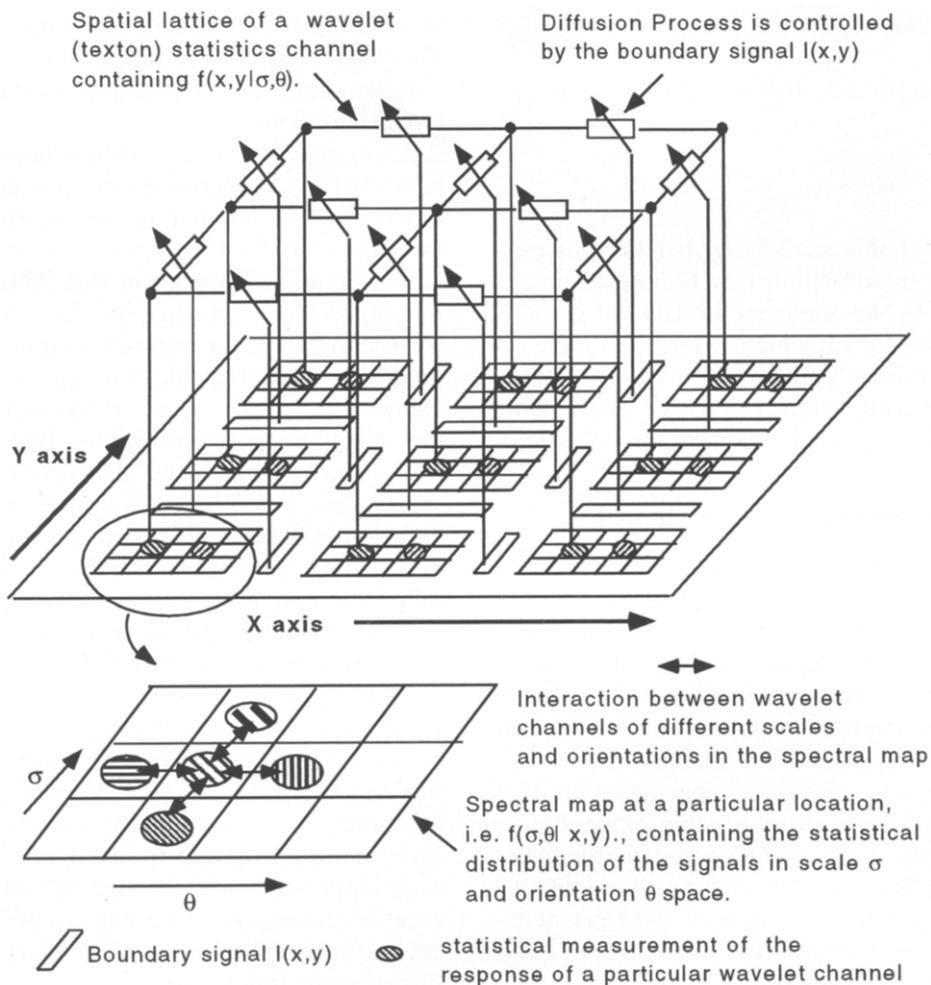


FIGURE 7. The architecture of computation of the Bayesian model: the 4D information is arranged as a spectral bundle. A set of 2D spectral maps are embedded in the 2D spatial space in a manner similar to hypercolumns embedded in a retinotopic coordinate in the primary visual cortex. Nodes of a wavelet channel are locally interconnected for spatial diffusion, and each channel is also connected to its proximal wavelet channels for spectral diffusion. The boundary signals can interrupt the spatial diffusion of the surface statistics signals.

strated the system's tolerance to gradual shifts in the scale or size of the texture elements.

The gradual sharpening of the boundary signals can be seen in Fig. 9, which shows the boundary signals of the system in response to the images Mondrain and Eye in successive relaxation stages. Initially, the boundary signal is fuzzy, fragmented and sketchy, resembling an artist's initial sketch. Over time, it converges to a sharp boundary.

The initial response of a channel depends on the frequency and orientation tuning of the channel. But as optimization progresses, the signals can be seen to diffuse within the channel across space as well as into other channels. This effect can be seen in the surface signals of the system in response to image Shell as shown in Fig. 10. At the end, each channel seems to carry a trace of the global impression of the image, indicating that the texture statistics within each surface region have become homogeneous.

In general, this system takes about 10 iteration stages and about 2 hr on the MASPAP parallel computer to converge. However, theoretically, each iteration stage can be performed in one 'dynamic relaxation' analog

neural network in milliseconds. Therefore, the 10κ stages can be completed within 100 msec in the brain.

The model has at least three major limitations: first, complete contour of the surface boundary is not guaranteed because there is no hard constraint or surface-region label to enforce complete partitioning of the different surface regions; second, the parameters of the surface priors, i.e. the variances of the statistical distributions, are uniformly set *a priori* for the whole visual scene. For a system that automatically perform texture segmentation for different realistic images, these parameters should be adaptively estimated within each region. This lack of adaptation of parameters according to local image statistics leads to the emergence of spurious boundary signals on the surface of the shell (Shell), as well as on the surface of the sand dune (Sierra). We have developed a more advanced model to address these problems (Zhu, Lee & Yuille, 1995).

PSYCHOLOGICAL IMPLICATION

This Bayesian model provides a useful framework for understanding the computation in the visual system that

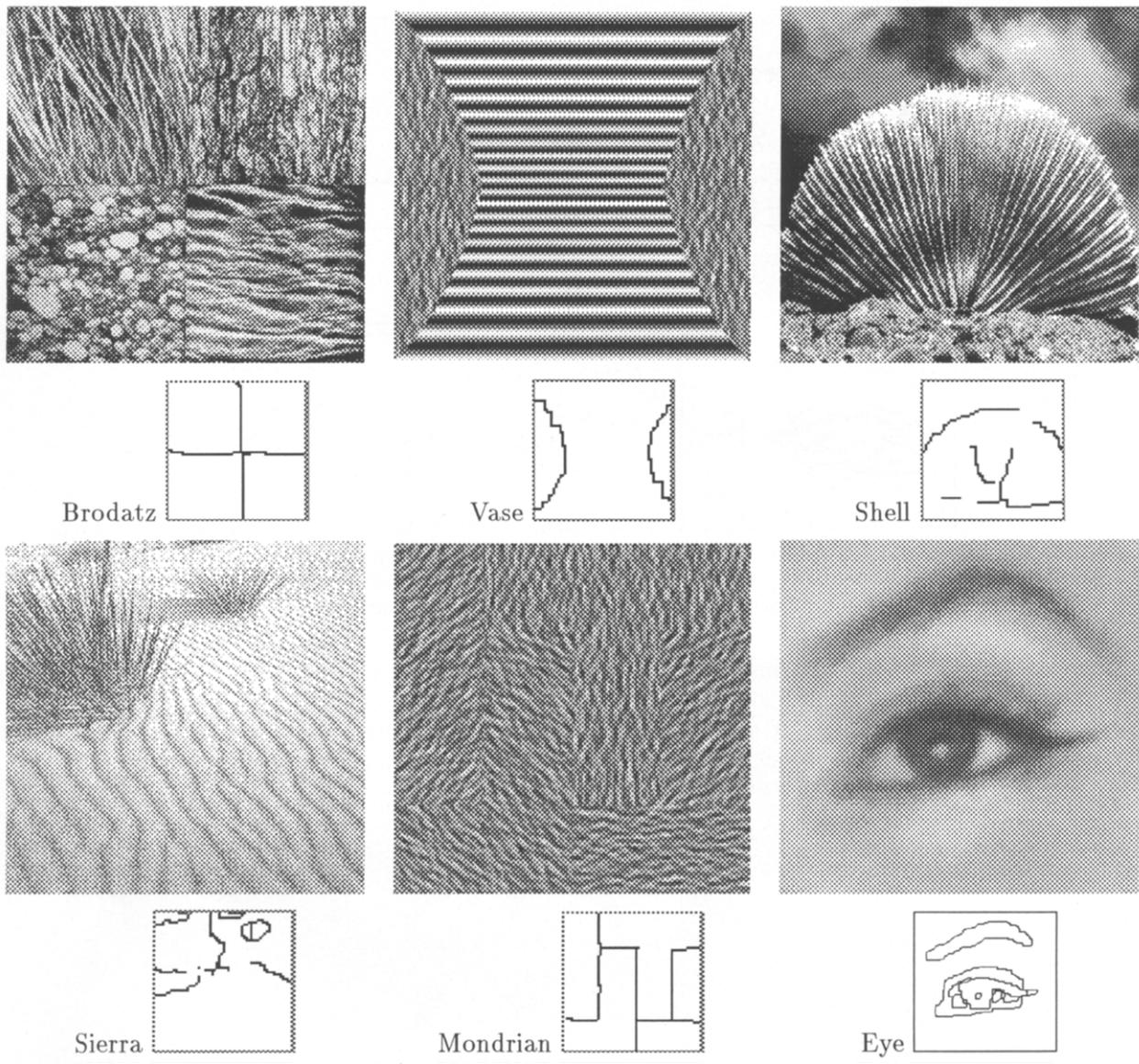


FIGURE 8. The boundary output of the Bayesian model in response to a set of texture images. Parameters used for Vase, Mondrian and Shell are $\lambda = 6$, $\alpha = 0.02$, $\gamma_s = 2$, $\gamma_\theta = 1$. Parameters for Brodatz and Sierra are $\lambda = 12$, $\alpha = 0.005$, $\gamma_s = 4$, $\gamma_\theta = 4$. For the image Eye, only the luminance channel is relevant, therefore segmentation can be achieved using a single weak-membrane with parameters $\lambda = 6$, $\alpha = 0.03$.

results in the perceptual experience of texture segmentation elucidated in the studies of Julesz (1983), Beck *et al.* (1983) and Treisman (1985). In this and the next sections, we attempt to build some conceptual links between the behaviors of the model and the psychological as well as the physiological observations.

First, psychologically, brightness, color and texture signals are found to be able to fill in artificial or natural scotoma to complete a surface (Yarbus, 1967; Lashley, 1941; Ramachandran, 1991; Paradiso & Nakayama, 1991; Redies & Spillman, 1981; Watanabe & Cavanagh, 1991). This psychological phenomenon of filling-in is related to the nonlinear diffusion for interpolating the surface statistical signals f in the texture segmentation process. However, the diffusion of surface statistical signals in our model is responsible for completing the percept of surface, rather than for creating the percept of interpolated texture pattern. The surface interpolation

process has a time course of 100–200 msec (Paradiso & Nakayama, 1991; Watanabe & Cavanagh, 1991), whereas the reconstruction of texture patterns to fill up the scotoma might take seconds (Lashley, 1941; Ramachandran, 1991). The texture filling-in process might have two components: surface interpolation and image reconstruction. Our model might capture the diffusive process of surface interpolation, but not the mechanism for perceptual image reconstruction.

Julesz (1983) has pointed out that during the 100 msec span of preattentive texture segmentation, the texton atoms have not been glued together to form a texton molecule; that the region grouping process seems to be operated by ‘averaging’ the density of texton atoms such as oriented segments. Because the averaging process is applied to the atoms of texture molecules, the preattentive system is blind to the positional and configurational information of textons, i.e. the structure of the molecules

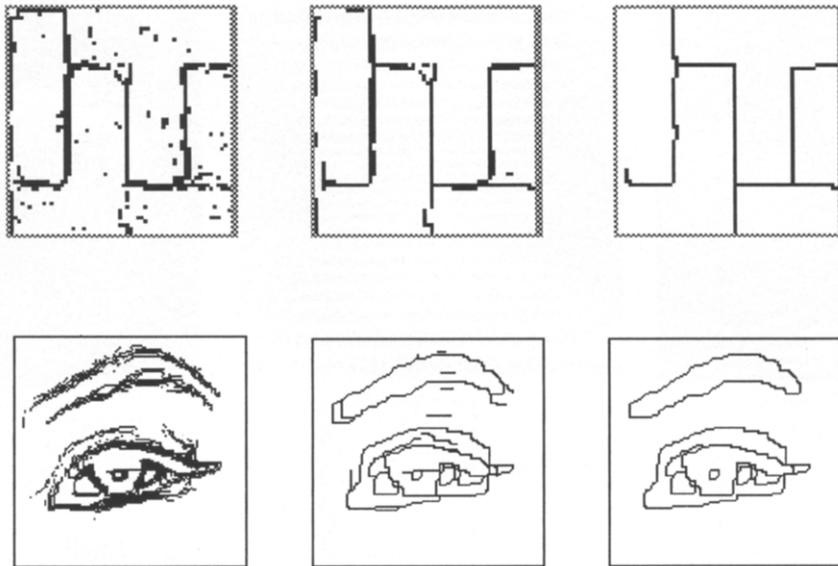
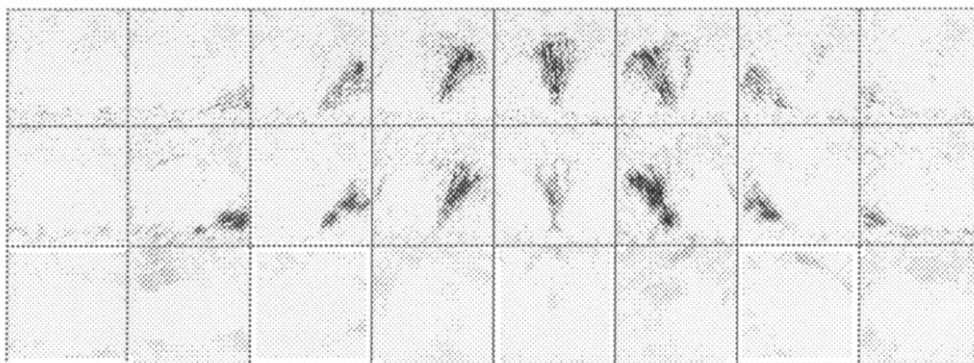


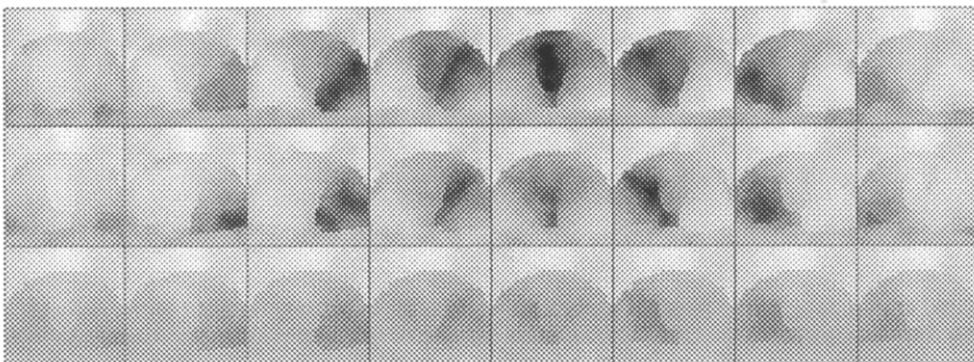
FIGURE 9. The evolution of the boundary signals over three successive κ ($\kappa = 0.5, \kappa \approx 0.1, \kappa = 0$) relaxation stages for Mondrian [top row] ($\alpha = 0.02, \lambda = 6, \gamma_a = 2, \gamma_\theta = 1$) and for Eye [bottom row] ($\alpha = 0.03, \lambda = 6$). The initial response resembles an artist's sketch, and the final response resembles a Matisse's drawing.

themselves. This phenomenon is consistent with the nonlinear spatial diffusion of the surface statistical signals f in the wavelet channels of our model. The insight provided by the model is that the 'averaging' process and the part of the 'filling-in' process are

both phenomena related to surface interpolation process but that the perceptual reconstruction of the filled-in patterns require additional processes in a longer time frame which are not captured in this model.



Initial response map of the surface statistical channels.



Final response map of the surface statistical channels.

FIGURE 10. The normalized surface statistics of the model in response to image Shell. Within each response map, there are 3×8 little squares, each showing the spatial response of a particular wavelet channel. The first column on the left is computed from horizontally tuned wavelets. Orientation selectivity rotates with the columns. The top row is the finest scale and scales increase with row. The initial response is very localized in spectral and spatial space. The final response shows significant diffusion in all dimensions, with the shell's unified surface cartoon appearing in every channel.

An unique component of our model is its explicit priors on the statistical variations in the scale and orientation domains within a surface region. These priors introduce a smoothing interaction between texton statistics channels of different scales and orientations. The magnitude of this smoothing interaction depends on the global texture statistics within a surface region. When a surface contains random or non-stationary texture, significant smoothing in scale and orientation space will occur to capture the variance of texton statistics. Such smoothing will reduce the difference of the texton statistical distributions of two random texture regions. In this case, segmentation requires a larger difference in scale or orientation in the textures of the two regions. Therefore, our model suggests that texture segmentation is not a *local* process, but is a context-sensitive *global process* involving surface interpolation. The global statistics of a surface determines the texture gradient threshold and the speed of texture segmentation. This is consistent with Enns (1986) and Nothdurft's (1991) findings that texture segmentation is context sensitive and that a larger texture gradient is required for segmenting regions of increasing random texture elements.

The model also provides an interesting insight to some common experiences. The gradual refinement of the boundary signals in the system is similar to the evolution from a quick sketch to a clean line drawing of an artist. People perceive cartoons, caricatures and line-drawings much faster than real images. In a cartoon, the outline of the figures resembles the output of the boundary system and the coloring of the cartoon figures resembles the output of the surface system. By discarding and simplifying the images, cartoons allow us to skip several steps in our visual processing, leading to more immediate perception of forms.

BIOLOGICAL IMPLICATION

We propose that texture segmentation, as a part of the surface inference process, is mostly accomplished in the primary visual cortex. This proposal is based on several pieces of evidence. First, surface perception can be induced preattentively by stereo information (Nakayama, Shimojo & Silverman, 1989). Since cells lose their specificity for stereo disparity and become binocular beyond V1, the creation of the surface percept likely takes place within V1. Second, the receptive fields of V1 cells resemble Julesz's textons. Therefore, the representation of textons and the computation of texton density and gradients likely locate in V1 so that the segmentation and pop-out information can be immediately feedback to the subcortical areas to direct reflexive eye movement. Third, even if the texton statistics are computed in extrastriate cortices, texture boundary would still be represented in V1 because only V1 cells have the high spatial specificity required for boundary localization. In fact, Knierim and Van Essen (1992), Lamme (1995), Zipser, Lee, Lamme and Schiller (1994), and Lee, Mumford and Schiller

(1995) have found ample evidence that V1 cells are sensitive to texture contrast.

Neurons in the primary visual cortex have been roughly classified into simple, complex and hypercomplex cells. Simple cells have high spatial specificity and behave more or less like linear wavelet filters. Complex cells are distinguished from simple cells by two fundamental properties: (1) they showed unmodulated responses to drifting sinusoidal gratings; (2) they are sensitive to contrast but insensitive to the direction of contrast. Hypersimple and hypercomplex cells are simple and complex cells with additional end-stopping or side-stopping properties. Although the receptive fields of simple cells can be modeled by Gabor wavelets and therefore can be understood as representational elements as discussed earlier, simple cells can also be understood as the first and the second Gaussian directional derivatives (Young, 1985) on the luminance signals. Complex cells can be considered as computing the statistics of the first and second derivatives of the luminance (simple cells' responses) and hypercomplex cells as taking the first and second derivatives of the complex cell responses. To push this idea one step further, we conjecture that the complex cells are not just computing surface feature statistics, but are also actively interpolating the surface, thus constituting the surface system. The hypercomplex cells are not just computing texture gradient, but are also detecting boundary as well, thus constituting the boundary system. In this view, simple cells, complex cells and hypercomplex cells can be understood within a unified framework: they are computational units for a boundary detection and surface interpolation system for luminance, textures and likely for other visual cues such as color, motion, stereo disparity as well.

Many of the well known context-sensitive phenomena of complex and hypercomplex cells, i.e. sensitivity of visual neurons of global features that fall outside their classical receptive fields, can be understood as behaviors of computational units in this Bayesian surface interpolation framework. The iso-orientation surround inhibition (Fries, Albus & Creutzfeldt, 1977; Nelson & Frost, 1978), side-stopping (De Valois, Thorell & Albrecht, 1985; Born & Tootell, 1991), cross-orientation surround facilitation or orientation contrast sensitivity (Gilbert & Wiesel, 1990; Knierim & Van Essen, 1992; Lamme, 1995) are suggestive of the texture contrast sensitivity and lateral-inhibition characteristic of the cells participating in boundary detection. On the other hand, complex cells that are not sensitive to the surround stimuli (Knierim & Van Essen, 1992) or are facilitated by surround stimuli composed of iso-orientation and iso-frequency features (Maffei & Fiorentini, 1976; Gilbert & Wiesel, 1990) are suggestive of the cells involving in surface interpolation. In fact, the unmodulated responses of complex cells to drifting sinusoidal gratings could arise simply from the diffusion within the surface statistics channel.

Some insights provided by the model have been confirmed by recent neurophysiological experiments. Lee *et al.* (1995) found that some complex cells'

responses to texture boundaries contract spatially over time, similar to the gradual sharpening of boundary signals predicted by the model [see equations (12) and (13)]; that the time required for the emergence of the sharp boundary signals is proportional to the area of the texture region, confirming that texture boundary detection is not a local process, but is coupled with a global surface interpolation mechanism. Moreover, Lamme (1995) and Zipser *et al.* (1994) found that V1 cells are sensitive simultaneously to many segmentation cues, thus providing a substrate for cue-invariant surface interpolation and boundary detection. Lastly, Pettet and Gilbert (1991), De Weerd, Gattass, Desimone and Ungerleider (1994) have found V1 cells in artificial scotoma could develop response to surrounding stimuli even when there is nothing falling on to their classical receptive fields. These could be the neural correlates of the filling-in effect related to surface interpolation.

Two additional predictions of our model about the surface interpolation mechanism remain to be tested: (1) complex cells performing surface interpolation will exhibit a spatial smoothing effect within a surface region, i.e. the spatial variation of their responses over a texture region will decrease over time; (2) they should also exhibit the spectral diffusion effect, i.e. their spatial frequency or orientation bandwidth will expand over time when the texture within the surface is random or non-stationary. This spectral bandwidth expansion is dependent on the statistics of the images. It is proportional to the variance of the texture statistics within the whole surface region. Because of this, traditional sinusoidal gratings will produce very little bandwidth expansion because their spectral content is very concentrated.

V1 complex cells are probably too pluralistic to be classified simply into surface cells and boundary cells. These two classes of cells might represent the two extremes of a continuum, i.e. there will be cells that perform a variable amount of both tasks. The cortex develops and maintains itself in a very flexible way. It can adopt whatever happens to be there to perform the necessary function. The circuitry that evolves to make use of a whole spectrum of cells to realize the surface interpolation function would necessarily be quite pluralistic as well.

POSSIBLE NEURAL CIRCUITS AND MECHANISMS

The basic circuitry for implementing this Bayesian model is relatively simple and biologically plausible. Since the surface interpolation process involves primarily local horizontal interaction within the complex cell channels, it can be mediated by the well known horizontal axonal collaterals of the pyramidal complex cells in layers II + III of V1. These collaterals enable complex cells to connect to other cells up to 4–6 mm (i.e. 4–6 hypercolumns) away (Rockland & Lund 1983). Cells of similar tuning in different hypercolumns are interconnected and these connections tend to be excitatory in nature (Ts'o, Gilbert & Wiesel, 1986). The fact that the

interaction is not limited to cells of the same orientation specificity but extends to cells of *similar* orientation preference provides a physical substrate for spectral diffusion. The simplest neural circuit for implementing equation (11) (surface signal propagation) and equation (12) (boundary signal contraction) is shown in Figure 11.

In this circuit, complex cells of similar specificities in adjacent hypercolumns are interconnected with excitatory connections to mediate the spreading of the surface signals f in four dimensions. The boundary complex cells $l(x, y)$, on the other hand, are activated by the 'texton density gradient' (∇f) computed within a scale-dependent local window. The interaction between surface complex cells can be interrupted by shunting inhibition from the boundary cells, i.e. the $(1 - l(x, y))$ term. This interruption can be done by shunting the corresponding dendritic trees of the surface complex cells via basket cells, or it can be done via an ensemble of inhibitory interneurons such as the double bouquet cells whose vertical axons form a dense, long and narrow 'curtain' from layer II to V isolating one cortical column from another (Somogyi & Cowey, 1981).

The lateral interaction within V1 can produce non-linear smoothing of the surface statistics signals within a surface region. A weak excitatory coupling is sufficient to produce local coherent oscillation or synchronized firings among surface cells within a surface (Wilson & Bower, 1991). Temporal synchronization of groups of neurons could label the surface of an object (Von der Malsburg & Schneider, 1986) and evidence of stimulus-specific synchronization of V1 neurons have emerged over the past few years (Gray & Singer, 1989). More global synchronization can be produced through global feedback of the extrastriate cortices—a feature that is not captured in our present model.

Since most part of the visual scene are surfaces, surface cells would tend to be activated more often than boundary cells, thus more strongly labeled by cytochrome oxidase. This leads to the hypothesis that the cytochrome oxidase (CO) rich blobs in V1 are populated by cells coding surface signals and the interblobs are populated by cells coding boundary signals. This hypothesis is consistent with Rockland and Lund's (1983) findings that the CO blobs are included in the densely connected lattices in the superficial layer of V1.

What is the mechanism for controlling the dynamics of successive gradual relaxation (i.e. the parameter κ) in this neural network?

The work of Geiger and Yuille (1991) has related the stage of gradual relaxation with the slope of the activation function of the boundary process using the mean field approximation of statistical physics,

$$l(x, \tilde{\kappa}) = \frac{1}{1 + \exp\{-z/\tilde{\kappa}\}} \quad (19)$$

where $z = [\lambda^2 \|\nabla \tilde{f}(x, \tilde{\kappa})\|^2 - \alpha]$. The continuation parameter $\tilde{\kappa}$ is proportional to κ in equation (13) and z is the input to the boundary cell.

Since $\tilde{\kappa}$ controls the slope of the activation curve, the optimization process could be mediated by a phasic

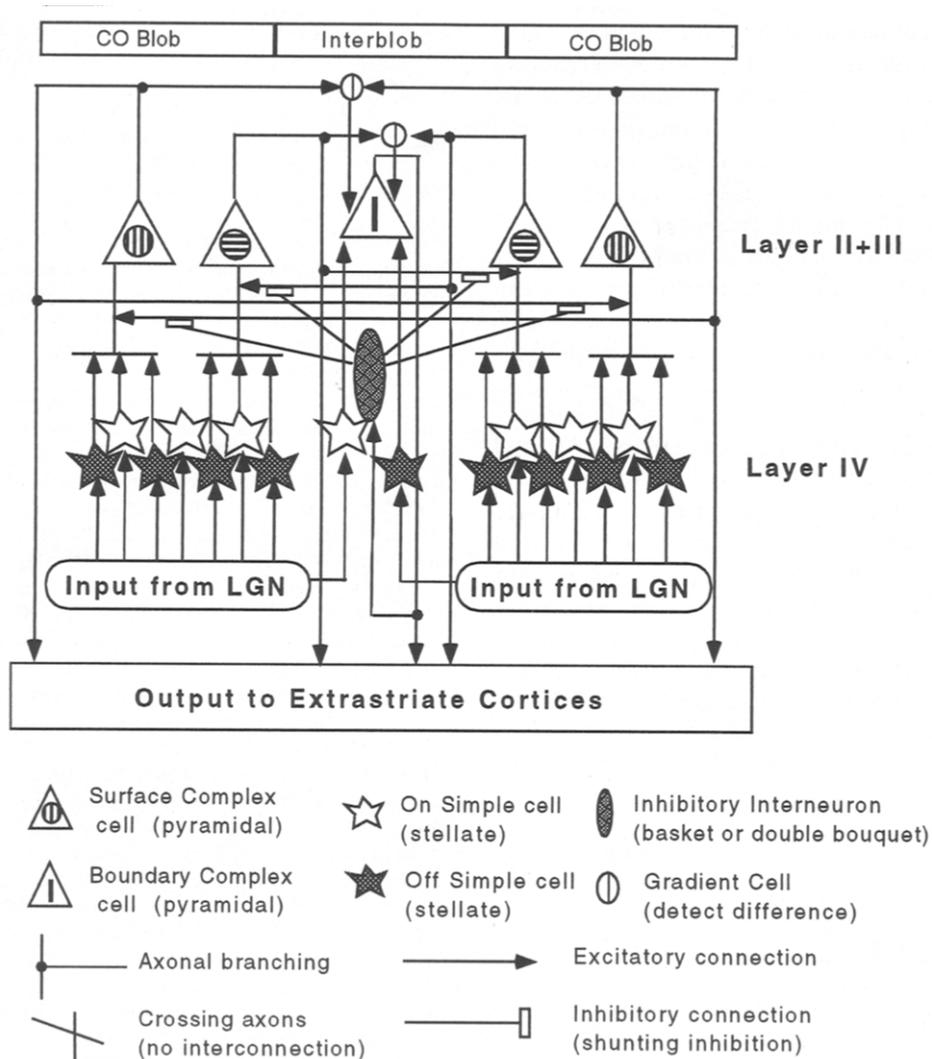


FIGURE 11. A hypothetical neural circuit for surface interpolation and boundary detection in V1: the LGN cells project to ON and OFF simple cells which then converge to complex cells. The surface complex cells are interconnected by the excitatory horizontal axonal collaterals. The gradient detectors compute gradients in the response of each channel, and they sum to activate the boundary complex cells which in turn can interrupt the mutual excitation of the surface cells by shunting inhibition.

increase in the boundary cell's sensitivity. During each optimization, as $\tilde{\kappa}$ moves toward 0, the slope of the sigmoidal activation function of the boundary cell will slowly increase, converging to a step function at the end. This gradual increase in discriminability, coupled with the diffusive expansion of the surface signals, produces the gradual sharpening of the boundary signals.

One plausible neural mechanism for implementing this kind of phasic increase in sensitivity is the PGO waves that have been observed to propagate from the brain stem to the LGN and the occipital cortex during and after each saccadic eye movement (Steriade, Pare, Hu & Deschenes, 1990). These cholinergic waves appear to produce a phasic increase in arousal level leading to excitability enhancement in the visual cortical cells (Livingston & Hubel, 1981). The duration of each burst of these PGO waves is about 150–350 msec, consistent with the 100 msec time-frame required for texture discrimination (Julesz, 1983) and the completion of brightness filling-in (Paradiso & Nakayama, 1991). During the 250 msec fixation period between two

saccades, there are about 10 bursts of spikes (Gray & Singer, 1989). We can envision each burst of spikes corresponds to one gradient descent at a relaxation stage. Thus, the optimization process involving about 10 relaxation stages can be completed during the fixation interval.

CONCLUSION

We suggested that texture segmentation is a part of the computational process for Bayesian surface inference. This computation can be described in a highly idealized sense as the minimization of a certain energy functional. The dynamical processes involved in the gradual relaxation algorithm used to minimize the functional are characterized by a gradual sharpening of the boundary signals and a simultaneous smoothing of the surface signals in a 4D spatial-spectral domain.

The relevance of this class of models to the visual system in solving the ill-posed early vision problems has been suggested by a number of workers (Poggio,

Torre & Koch, 1985; Koch, Marroquin & Yuille, 1985). Neural network models of similar favor have also been proposed by Grossberg and Mingola (1985). By introducing the proper psychological and physiological constraints, we establish a tighter link between the computational theory, neural network models and biological reality. The model provides a functional framework for understanding complex cells as boundary cells and surface cells in a surface interpolation system. The model also provides specific predictions on boundary and surface complex cells. Some of these predictions have been confirmed.

We must recognize that the energy functional discussed in this paper is a highly simplified and idealized description of reality. The actual working of the brain, even at the level of preattentive vision, is far more complicated than any simple energy functional can possibly represent. Despite its limitations, the present simple model is illustrative of the basic characteristics of the system and can provide some basic insights to the processes of texture segmentation and surface interpolation in the primary visual cortex.

REFERENCES

- Albrecht, D. G. & Geisler, W. S. (1991). Motion sensitivity and the contrast-response function of simple cells in the visual cortex. *Visual Neuroscience*, 7, 531-546.
- Ambrosio, L. & Tortorelli, V. M. (1990). On the approximation of free discontinuity problems. *Preprints di Matematica*, n 86. Pisa, Italy: Scuola Normale Superiore.
- Beck, J., Prazdy, K. & Rosenfeld, A. (1983). A theory of texture segmentation. In Beck J., Hope B. & Rosenfeld A. (eds), *Human and machine vision*. New York: Academic Press.
- Blake, A. & Zisserman, A. (1987). *Visual reconstruction*. Cambridge, Mass: MIT Press.
- Blasdel, G. G. (1989). Topography of visual function as shown with voltage-sensitive dyes. In Lund J. S. (Ed.), *Sensory processing in the mammalian brain* (pp. 242-267). Oxford: Oxford Univ. Press.
- Born, R. T. & Tootell, R. B. H. (1991). Single-unit and 2-deoxyglucose studies of side inhibition in macaque striate cortex *PNAS*, 88, 7071-7075.
- Bovik, A. C., Clark, M. & Geisler, W. S. (1990). Multichannel texture analysis using localized spatial filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12, January.
- Daugman, J. G. (1985). Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *Journal of the Optical Society of America*, 2, 1160-1169.
- De Valois, R. L. & De Valois, K. K. (1988). *Spatial vision*. New York: Oxford University Press.
- De Valois, R. L., Thorell, L. G. & Albrecht, D. G. (1985). Periodicity of striate-cortex-cell receptive fields. *Journal of the Optical Society of America*, A2, 1115-1123.
- De Weerd, P., Gattass, R., Desimone, R. & Ungerleider, L. G. (1994). Neural mechanisms for perceptual filling-in. *Society for Neuroscience Abstracts*, 20, 1053.
- Enns, J. (1986). Seeing textons in context. *Perception & Psychophysics*, 39, 143-147.
- Fogel, I. & Sagi, D. (1989). Gabor filters as texture discriminator. *Biological Cybernetics*, 61, 103-113.
- Fries, W., Albus, K. & Creutzfeldt, O. D. (1977). Effects of interacting visual patterns on single cell responses in cat's striate cortex. *Vision Research*, 17, 1001-1008.
- Geiger, D. & Yuille, A. (1991). A common framework for image segmentation. *International Journal of Computer Vision*, 6, 227-243.
- Geman, D., Geman, S., Graffigne, C. & Dong, P. (1990). Boundary detection by constrained optimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12, 609-628.
- Gilbert, C. D. & Wiesel, T. N. (1990). The influence of contextual stimuli on the orientation selectivity of cells in primary visual cortex of the cat. *Vision Research*, 30, 1689-1701.
- Gray, C. M. & Singer, W. (1989). Stimulus specific neuronal oscillations in orientation columns of cat visual cortex. *Proceedings of the National Academy of Science, U.S.A.*, 86, 1698-1702.
- Grossberg, S. & Mingolla, E. (1985). Neural dynamics of perceptual grouping: Textures, boundaries, and emergent segmentations. *Perception & Psychophysics*, 38, 141-171.
- Heeger, D. (1992). Normalization of cell responses in cat striate cortex. *Visual Neuroscience*, 9, 181-197.
- Hubel, D. H. & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology London*, 195, 215-243.
- Julesz, B. (1983). Toward an axiomatic theory of preattentive vision. In Edelman, G. M., Einar Gall, W. & Maxwell Cowan, W. (Eds), *Dynamic aspects of neocortical function*. New York: Wiley.
- Julesz, B. (1986). Texton gradients: The texton theory revisited. *Biological Cybernetics*, 22, 107-119.
- Knierim, J. J. & Van Essen, D. (1992). Neuronal responses to static texture patterns in area V1 of the alert macaque monkey. *Journal of Neurophysiology*, 67, 961-980.
- Koch, C., Marroquin, J. & Yuille, A. L. (1986). Analog "neuronal" networks in early vision. *Proceedings of the National Academy of Science, U.S.A.*, 83, 4263-4267.
- Lamme, V. A. F. (1995). The neurophysiology of figure-ground segregation in primary visual cortex. *Journal of Neuroscience*, 15, 1605-1615.
- Lashley, K. S. (1941). *Archives of Neurology & Psychiatry*, 46, 331-339.
- Lee, T. S. (1994). Image representation using 2D Gabor-wavelets. Accepted in *IEEE Transactions on Pattern Analysis and Machine Intelligence*. Short version published in *Proceedings of 1994 International Conference in Image Processing* (Vol. 2, 590-594).
- Lee, T. S., Mumford, D. & Yuille, A. L. (1992). Texture segmentation by minimizing vector-valued energy functionals: The coupled-membrane model. In Sandini G. (Ed.), *Lecture Notes in Computer Science*, 588, Computer Vision ECCV'92 (pp. 165-173). Berlin: Springer.
- Lee, T. S., Mumford, D. & Schiller, P. H. (1995). Neural correlates of boundary and medial axis representation in V1. *Investigative Ophthalmology & Visual Science*, 36, 2205.
- Livingston, M. & Hubel, D. H. (1981). Effects of sleep and arousal on the processing of visual information in cat. *Nature*, 291, 554-561.
- Livingston, M. & Hubel, D. H. (1984). Specificity of intrinsic connections in primate primary visual cortex. *Journal of Neuroscience*, 4, 2830-2835.
- Maffei, L. & Fiorentini, A. (1976). The unresponsive regions of visual cortical receptive fields. *Vision Research*, 16, 1131-1139.
- Malik, J. & Perona, P. (1990). Preattentive texture discrimination with early vision mechanisms. *Journal of the Optical Society of America A*, 7, 923-932.
- Nakayama, K., Shimojo, S., Silverman, G. H. (1989). Stereoscopic depth: Its relation to image segmentation, grouping, and the recognition of occluded objects. *Perception*, 18, 55-68.
- Nelson, J. I. & Frost, B. (1978). Orientation selective inhibition from beyond the classical visual receptive field. *Brain Research*, 139, 359-365.
- Nothdurft, H. C. (1991). Texture segmentation and pop-out from orientation contrast. *Vision Research*, 31, 1073-1078.
- Paradiso, M. A. & Nakayama, K. (1991). Brightness perception and filling-in. *Vision Research*, 31, 1221-1236.
- Pettet, M. W. & Gilbert, C. D. (1992). Dynamics changes in receptive field size in cat primary visual cortex. *Proceedings of the National Academy of Science, U.S.A.*, 89, 8366-8370.
- Pollen, D. A., Gaska, J. P. & Jacobson, L. D. (1989). Physiological constraints on models of visual cortical function. In Cotterill, R. M. J. (Ed.), *Models of brain function*. England: Cambridge University Press.

- Poggio, T., Torre, V. & Koch, C. (1985). Computational vision and regularization theory. *Nature*, 317, 26 September.
- Ramachandran, V. S. (1991). Perceptual filling in of artificially induced scotomas in human vision. *Nature*, 350, 26 September.
- Redies, C. & Spillman, L. (1981). The neon color effect in the Ehrenstein illusion. *Perception*, 10, 667–681.
- Richardson, T. J. (1990). Scale independent piecewise smooth segmentation of images via variational methods. Ph.D. thesis, MIT.
- Rockland, K. S. & Lund, J. S. (1983). Intrinsic laminar lattice connections in primate visual cortex. *Journal of Comparative Neurology*, 216, 303–318.
- Schiller, P. H., Finlay, B. L. & Volman, S. F. (1976). Quantitative studies of single-cell properties in monkey striate cortex. I. Spatiotemporal organization of receptive fields. *Journal of Neurophysiology*, 39, 1288–1319.
- Somogyi, P. & Cowey, A. (1981). Combined Golgi and electron microscopic study on the synapses formed by double bouquet cells in the visual cortex of the cat and monkey. *Journal of Comparative Neurology*, 195, 547–566.
- Steriade, M., Pare, D., Hu, B. & Deschenes, M. (1990). The visual thalamocortical system and its modulation by the brain stem core. *Progress in sensory physiology*, 10. Berlin: Springer.
- Treisman, A. (1985). Preattentive processing in vision. *Computer Vision, Graphics and Image Processing*, 31, 156–177.
- Ts'o, D., Gilbert, C. D. & Wiesel, T. N. (1986). Relationships between horizontal interactions and functional architecture in cat striate cortex as revealed by cross-correlation analysis. *Journal of Neuroscience*, 6, 1160–1170.
- Turner, M. R. (1986). Texture discrimination by Gabor functions. *Biological Cybernetics*, 55, 71–82.
- Von der Malsburg, C. & Schneider, W. (1986). A neural cocktail-party processor. *Biological Cybernetics*, 54, 29–40.
- Voorhees, H. & Poggio, T. (1988). Computing texture boundaries from images. *Nature (London)*, 333, 364–367.
- Watanabe, T. & Cavanagh, P. (1991). Texture and motion spreading and transparency. Submitted to *Perception and Psychophysics*.
- Wilson, M. A. & Bower, J. M. (1991). A computer simulation of oscillatory behavior in primary visual cortex. *Neural Computation*, 3, 498–509.
- Yarbus, A. L. (1967). *Eye movements and vision*. New York: Plenum Press.
- Young, R. A. (1985). The Gaussian derivative theory of spatial vision: Analysis of cortical cell receptive field line-weighting profiles. General Motors Research Technical Report, GMR-4920.
- Zipser, K., Lee, T. S., Lamme, V. & Schiller, P. H. (1994). Neural correlates of cue-invariant scene segmentation. *Investigative Ophthalmology & Visual Science*, 35, 1973.
- Zhu, S. C., Lee, T. S. & Yuille, A. (1995). Region competition: Unifying snakes, region growing and Bayes/MDL for image segmentation. *Proceedings of 1995 International Conference of Computer Vision*.

Acknowledgements—This research is supported in part by a Harvard-MIT Division of Health Science and Technology fellowship and a Harvard University DAS fellowship to the author and a NSF Frant DMS 91-21266 to David Mumford. The author is indebted to David Mumford, Alan Yuille, John Daugman, Richard Kronauer, Danial Pollen, Richard Born, Gary Blasel, Roger Tootell, Peter Hallinan, Peter Belhumeur, Song-Chun Zhu, Liu Zheng, Michael Weisman, I.-Han Chou, Marc Sommer, Edward Tehovnik and Karl Zipser for interesting discussions and valuable technical assistance.