Texture Segmentation by Minimizing Vector-Valued Energy Functionals: The Coupled-Membrane Model *

Tai Sing Lee, David Mumford, Alan Yuille

Harvard Robotics Laboratory,
Division of Applied Sciences, Harvard University, Cambridge, MA 02138.

Abstract. This paper presents a computational model that segments images based on the textural properties of object surfaces. The proposed Coupled-Membrane model applies the weak membrane approach to an image $\mathcal{W}I(\sigma, \theta, x, y)$, derived from the power responses of a family of self-similar quadrature Gabor wavelets. While segmentation breaks are allowed in x and y only, coupling is introduced to in all 4 dimensions. The resulting spatial and spectral diffusion prevents minor variations in local textures from producing segmentation boundaries. Experiments showed that the model is adequate in segmenting a class of synthetic and natural texture images.

1 Introduction

This paper presents a computational model that segments images based on the textural properties of object surfaces. The proposed model distinguishes itself from the previous models in texture segmentation [Turner 1986, Voorhees and Poggio 1988, Malik and Perona 1989, Fogel and Sagi 1989, Bovik, Clark and Geisler 1990, Reed and Wechsler 1990, Geman et al 1990] in the following way.

Previous models have started with the extraction from the image $I(x, y)$ of some set of texture features which can be viewed as forming auxiliary texture images $I_\alpha(x, y)$. Then applying either region growing, boundary detection, or (in the single paper [Geman et al 1990]) a membrane-like method combining these two, a segmentation is derived. In our model, the texture features are the power responses of quadrature Gabor filters. These filters form a continuous family depending on two variables σ, θ, and can be derived like wavelets from dilation and rotation of a single filter. Thus we think of the texture features as combining into a single image $\mathcal{W}I(\sigma, \theta, x, y)$ depending on *4 continuous variables*. We apply the weak membrane approach to segmenting this signal, in which coupling is introduced in all 4 dimensions, but breaks are allowed in x and y only. We call this the *Coupled- Membrane model*.

Why is this model useful? Previous methods generally deal only with textures that are statistically stationary (i.e. approximately translationally invariant) and not too granular (e.g. with widely spaced textons, or large local variations). But natural textures do not satisfy either: Firstly, they show considerable texture 'gradients', in which the power distribution of the texture among various channels changes slowly but systematically over

a region, due for instance to the perspective affine distortion imposed on surface features of solid objects in a 3-dimensional world, and to the deformation caused by the non-planarity of the objects' body shapes. J. J. Gibson [Gibson 1979] has emphasized how these texture gradients are ubiquitous clues to the 3D structure of the world. Secondly, they show random local fluctuations, due to the stochasticity in their generation processes, which are often quite large (compare the four subimages in 'Mosaic' below, taken from [Brodatz 1966]). The inter-membrane couplings unique to our Coupled-Membrane model allow interaction between neighboring components in the spectral vector and prevent minor local variations from producing segmentation boundaries. At the same time, they introduce explicitly the appropriate metric between texture channels so that a shift in the peak of the power spectrum to a nearby frequency or orientation is treated differently from a shift to a distant frequency or orientation. As we shall see, this allows us to begin to solve these problems for natural textures.

This paper is organized as follows: first, we will discuss how texture is represented in our model and how texture disparity can be computed from this representation. Then, we will discuss the Coupled-Membrane model for texture segmentation in its continuous formulation and discrete approximation. Finally, we will present our experimental results.

2 Gabor-Wavelet Representation of Texture

Texture segmentation requires a description of local texture properties in an image. Previous methods include texton statistics [Voorhees and Poggio 1988], DOG filters [Malik and Perona 1989], windowed Fourier transform or Gabor filtering [Turner 1986, Fogel and Sagi 1989, Bovik, Clark and Geisler 1990, Reed and Wechsler 1990]. While the first two of these methods emphasize feature detection, the Gabor-Fourier method is based on power spectrum analysis or autocorrelation.

In our model, the texture features are the power responses of quadrature Gabor filters. These filters form a continuous family depending on two variables σ, θ, and can be derived like wavelets from dilation and rotation of a single filter. We call this *Gabor-Wavelet Representation*. Physiological evidence suggests that the visual cortex is employing a similar representation for encoding visual information. We impose the constraints derived from physiological data [Pollen et al. 1989, Daugman 1985] and obtain the following family of self-similar Gabor filters centered at $(x = 0, y = 0)$ in the spatial domain. (For details, readers are referred to our technical report [Lee, et al 1991].)

$$G(\sigma, \theta, x, y) = \frac{\sigma^2}{50\pi} e^{-\frac{\sigma^2}{100}(4(x\cos\theta + y\sin\theta)^2 + (-x\sin\theta + y\cos\theta)^2} \cdot e^{i(\sigma\cos\theta x + \sigma\sin\theta y)} \qquad (1)$$

where σ is the radial frequency, and θ is the angular orientation of the filter.

In a manner completely analogous to the generation of wavelet bases from a single basic wavelet, this whole family of Gabor filters can be generated by rotation and dilation from the following single Gabor filter (as shown in figure 1):

$$\phi(x, y) = \frac{1}{50\pi} e^{-\frac{1}{100}(4x^2 + y^2)} \cdot e^{ix} \qquad (2)$$

Self-similar Gabor filters from this family serve both as band-pass filters and multi-scale matched filters, producing a representational scheme that unifies power-spectrum analysis and feature detection.

The convolution of this family of filters with the image produces a single image $\mathcal{W}I(\sigma, \theta, x, y)$ which is the normalized power modulus of the filter ensemble as follows,
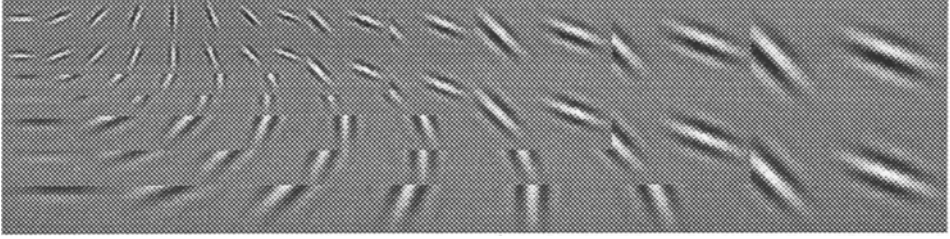
Fig. 1. Rotated and Dilated quadrature Gabor filters

$$\mathcal{W}I(\sigma, \theta, x, y) = \frac{1}{Z}|(G_{\sigma, \theta} * I)(x, y)|^2 \tag{3}$$

where

$$(G_{\sigma, \theta} * I)(x_o, y_o) = \int_x \int_y G_{\sigma, \theta}(x - x_o, y - y_o) I(x, y) dx dy \tag{4}$$

and

$$Z = \int \int |(G_{\sigma, \theta} * I)(x, y)|^2 d\sigma d\theta \tag{5}$$

Since each Gabor filter has a Gaussian spread in its frequency plane, local power spectrum of an image can be sampled in a parsimonious and discrete manner. In our implementation, we use Gabor wavelets with a sampling interval of 1-octave in frequency and 22.5^o in orientation to pave the spatial frequency plane, as shown in figure 2.
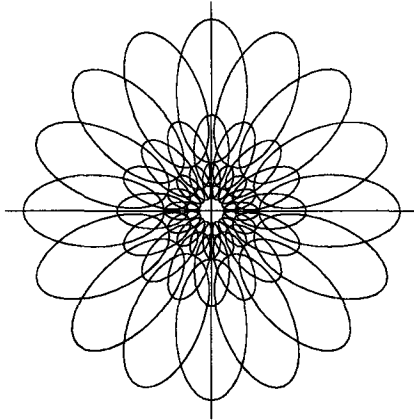

Fig. 2. Tiling of Spatial Frequency Plane by Gabor Wavelets

3 Texture Disparity and Spectral Proximity

Using this sampling scheme, we construct a spectral signature vector of 24 components (3 frequencies and 8 orientations) at each point (x, y) in the spatial domain. This vector is then normalized, i.e. divided by Z, to discount the luminance effect. As a result, each particular texture corresponds to a unit vector within this unit 24-dimensional ball. we advocate the use of the L_2 norm as the appropriate metric to compute the distance between two spectral signature vectors for the following reason.

The L_2 norm is superior to the L_∞ norm in computing texture disparity because it does not discard the proximity information between the spectral signatures of two texture patterns. When a L_∞ norm is computed, components in a spectral vector are treated as independent and their spectral proximity relationships are ignored. For instance, the L_∞ norm induced by rotating a texture by 30° will be the same as that induced by a 90° rotational shift. This is true independent of the sampling scheme. Although the L_2 norm behaves the same way as the L_∞ norm in a minimal sampling scheme with orthogonal bases, its value decreases for spectrally proximal textures when the bases become increasingly nonorthogonal in an oversampling scheme. In this case, the L_2 norm induced by a 30° rotation is smaller than that induced by a 90° rotation, as illustrated in figure 3.

Because the parsimonious scheme we used is not straightly a minimal one with orthogonal bases, it is benefited from the use the L_2 norm. The visual cortex, however, oversamples the power spectrum by at least two or three times as much in both σ and θ dimensions [Webster & De Valois 1985, Silverman et al 1989, Hubel and Wiesel 1977]. The proximity effect due to the L_2 norm would therefore be even more pronounced.

The parsimonious scheme saves computational effort, but also decreases the proximity effect. To compensate, we introduce smoothing in the spectral domain by coupling together the spectrally proximal components in the spectral vector, as will be discussed in the next section.
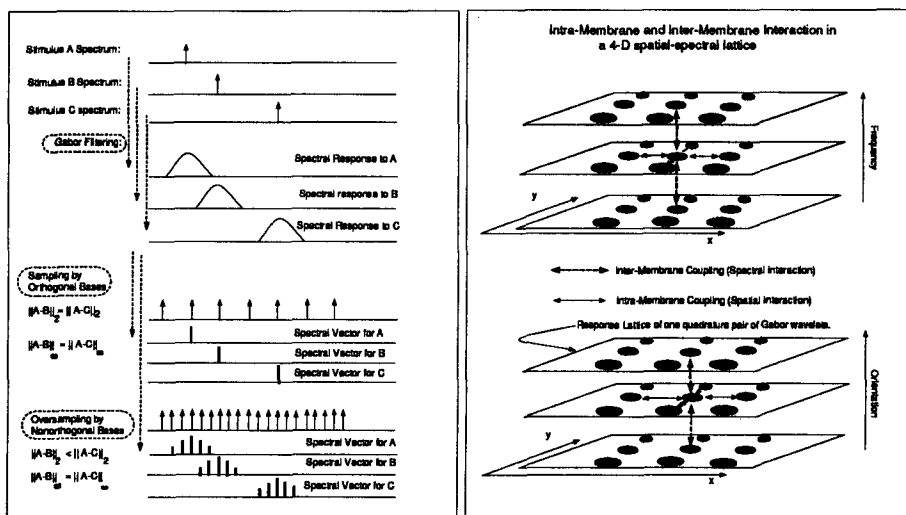

Fig. 3. Effect of L_2 norm.

Fig. 4. Couplings in the 4-D Lattice

4 Energy Functional for Texture Segmentation

The Coupled-Membrane model we developed for texture segmentation is a generalization of the Weak-Membrane Model [Blake and Zisserman 1987, Marroquin 1984, and Mumford and Shah 1985] or equivalently the Markov Random Field model [Geman and Geman 1984]. While the Weak-Membrane deals with intensity values in a 2-D image plane, our model deals with spectral responses in a 4-D spatial-spectral domain. The continuous formulation of the model is defined as follows,

Given the spectral signature image $\mathcal{W}I(\sigma, \theta, x, y)$, we are to find a piece-wise continuous functional $f(\sigma, \theta, x, y)$ that is its smooth estimation, with its texture noise and variations removed. Within a texture region, $f(\sigma, \theta, x, y)$ is continuous. Discontinuity in $f(\sigma, \theta, x, y)$ is allowed at the boundary in spatial domain between two texture regions.

These objectives are captured by following energy functional that is to be minimized,

$$
E(f, B) = \int \int_R \int \int_S \|f(\sigma, \theta, x, y) - \mathcal{W}I(\sigma, \theta, x, y)\|^2 d\log \sigma d\theta dx dy
$$
$$
+ \int \int_{R-B} \int \int_S [(\gamma_\sigma \frac{\partial f}{\partial(\log \sigma)})^2 + (\gamma_\theta \frac{\partial f}{\partial \theta})^2 + (\lambda \frac{\partial f}{\partial x})^2 + (\lambda \frac{\partial f}{\partial y})^2] d(\log \sigma) d\theta dx dy
$$
$$
+ \alpha \int_B ds
$$

where R and S are the finite 2-dimensional spatial and spectral domains respectively; boundaries $B \subset R$ is a finite set of piece-wise C^1 contours which meet ∂R and meet each other only at their endpoints. The contours of B cut R into a a finite set of disjoint regions $R_1, ..., R_n$, the connected components of R-B. The integration over S is done with $(d\log \sigma)d\theta = \frac{d\sigma d\theta}{\sigma}$ because the power spectrum is represented in log-polar form.

The first term of the energy functional forces the smoothed spectral response $f(\sigma, \theta, x, y)$ to be as close as possible to the measured spectral response $\mathcal{W}I(\sigma, \theta, x, y)$. The second term asks the spectral response to be as smooth as possible in both spatial and spectral domains. These two potentially antagonistic demands are to arrive at a compromise that is determined by the λ, γ_σ and γ_θ.

Since $f(\sigma, \theta, x, y)$ is required to be smooth only within each R_i but not across B, the third integral term is needed to prevent breaks from appearing everywhere. This term imposes a penalty α against each break and provides the binding force within a region.

5 Computer Implementation

To solve a functional minimization problem computationally, the energy functional is discretized as follows,

$$
E(f, B) = \sum_{i,j,k,l} (f(i, j, k, l) - \mathcal{W}I(i, j, k, l))^2 + \gamma_\sigma^2 \sum_{i,j,k,l} [f(i, j, k, l) - f(i, j, k, l+1)]^2
$$
$$
+ \gamma_\theta^2 \sum_{i,j,k,l} [f(i, j, k, l) - f(i, j, k+1, l)]^2
$$
$$
+ \lambda^2 \sum_{i,j,k,l} [f(i, j, k, l) - f(i+1, j, k, l)]^2 (1 - v(i + \frac{1}{2}, j))
$$

$$+\lambda^2 \sum_{i,j,k,l} [f(i,j,k,l) - f(i,j+1,k,l)]^2 (1 - h(i,j+\frac{1}{2}))$$

$$+\sum_{i,j}[v(i+\frac{1}{2},j) + h(i,j+\frac{1}{2})]$$

where i, j, k, l are indexes for x,y,θ, log σ respectively in the 4-dimensional spatial-spectral sampling lattice. v and h are vertical and horizontal breaks between the lattice points in the spatial domain.

Figure 4 illustrates the couplings among the nodes in the 4-D sampling lattice: Each membrane corresponds to $\mathcal{W}I(k,l)$ for a frequency l, and an orientation k. Within each membrane, each node is coupled to the nearest 4 neighboring nodes. At each spatial location, a membrane is coupled with 4 other membranes which are its nearest spectral neighbors.

As the segmentation-diffusion process unfolds, spectral response is allowed to diffuse from one node to its 4 spatial and 4 spectral nearest neighbors. Breaks, however, can only occur in the spatial domain. When the L_2 norm of the evolving membranes exceeds the texture disparity threshold $\sqrt{\frac{\alpha}{\lambda}}$ at a spatial location, a break will occur at that location to cut across all the membranes.

Given a set of values for parameters $\lambda, \gamma_\sigma, \gamma_\theta$, and α, an optimal compromise among the three terms in the energy functional produces a set of segmentation boundaries and smoothed spectral responses. Because the energy functional has many local minima due to its nonconvexity, the global optimal compromise can be obtained using special mathematical programming methods. This paper presents results obtained using a stochastic method called Simulated Annealing [Kirkpatrick 1983], and a deterministic method called Graduated Non-Convexity [Blake and Zeisserman 1987]. We implemented both two methods on DEC5000 workstation and on a massively parallel computer called MASPAR.

6 Experimental Results

A class of texture images, 256 x 256 pixels in size, are used to test the model. Perceptual boundaries in these images are defined primarily by difference in textures, and not by luminance contrast. The segmentation-diffusion is performed on a 64 x 64 spatial sampling grid. We use a simple annealing schedule schedule for Simulated Annealing: $T_n = 0.985T_{n-1}$ at each temperature step, with a starting temperature of 25. It takes about 24 hours on DEC 5000 or 6 hours on MASPAR to process each image. For GNC, the error resolution δ needs to be 2^{-12} to ensure the solution is close to the optimal one. It takes 140 hours on DEC 5000 or 7 hours on MASPAR. Despite the fast annealing schedule, the Simulated Annealing performs reasonably well. The answer provided by GNC, however, is closer to the global minimum. These algorithms have also been implemented in 1-D so that their solutions can be compared with the exact optimal solution yielded by dynamic programming.

Three images are presented here as illustrations: 'Vase' (figure 5), 'Mondrian' (figure 7a), and 'Mosaic' (figure 7b). 'Vase' is used to demonstrate the model's tolerance to texture 'gradient' due to inter-membrane coupling. When this coupling is disabled, the segmentation is not perceptually valid (figure 5d). The initial response and the final response of the filters to 'Vase' (figure 6) demonstrate the diffusion effect in both the spatial and spectral domains.

'Mondrian' and 'Mosaic' both demonstrate that the model's ability in segmenting synthetic and natural textures while withstanding significant texture noise, and local
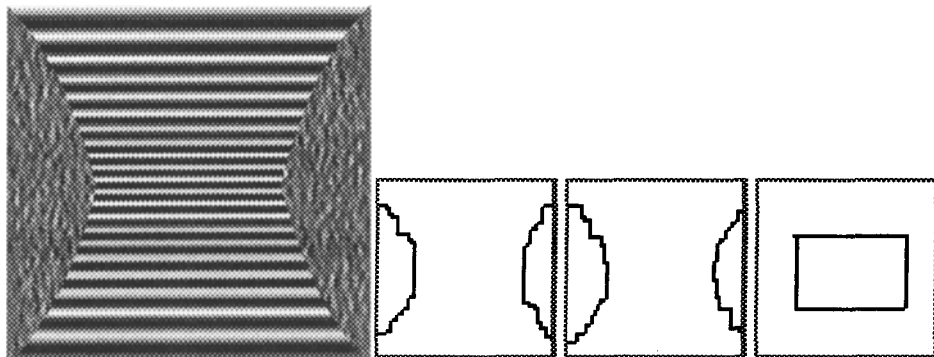
Fig. 5. (a) 'Vase' and its segmentations: (b) Simulated Annealing result with $\alpha = 0.02, \lambda = 6, \gamma_\theta = 2, \gamma_\sigma = 4$; (c) GNC result with $\alpha = 0.02, \lambda = 6, \gamma_\theta = 2, \gamma_\sigma = 4$; (d) GNC with $\alpha = 1.25, \lambda = 6, \gamma_\theta = \gamma_\sigma = 0$ i.e. without inter-membrane coupling.
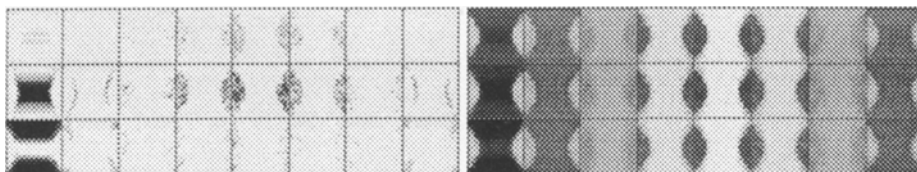

Fig. 6. (a) Initial filter response map for 'Vase'. (b) Final filter response map at the end of the segmentation-diffusion process (figure 5c). Each small square is the response map of a particular filter to the image. The maps are arranged in frequency rows (three frequencies) and orientation columns (eight orientations).

variation in scale and orientation. The initial and final response maps of 'Mondrian' (figure 8) underscore the cooperative effect of the diffusion and segmentation processes in producing sharp texture boundary from fuzzy input.

The parameter values used for the segmentation-diffusion process are shown in the figure caption. For the series of images we tested, the values needed to produce a segmentation similar to our perception are fairly close together.

7 Discussion:

The Coupled-Membrane Model with the Gabor-wavelet representation has produced promising results in the segmentation of a class of texture images. It combines the several sequential steps of filtering, smoothing and boundary detection in the previous texture segmentation models into a coherent and unified framework with a simple and elegant formalism. The model requires only three parameters (as γ_σ and γ_θ are related) and is more parsimonious in many aspects than the model Geman et al [1991] proposed. The issue of spectral proximity, ignored by the previous models, is addressed in our model by the introduction of spectral smoothing and the use of the L_2 norm with oversampling.

The model needs to be further developed to address to a wider class of natural images. In the form presented in this paper, the model has difficulty at the boundary between non-texture regions. This problem can be solved by incorporating into the model the
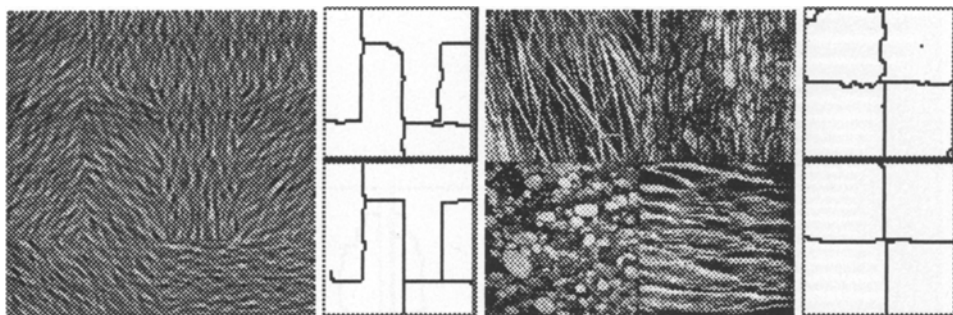
Fig. 7. (a) 'Mondrian' and its segmentation. Parameters: $\alpha = 0.02, \lambda = 6, \gamma_\theta = 2, \gamma_\sigma = 4$. (b) 'Mosaic' and its segmentation. Parameters: $\alpha = 0.02, \lambda = 12, \gamma_\theta = 2, \gamma_\sigma = 4$. Top segmentation: SA. Bottom segmentation: GNC.

Fig. 8. (a) Initial filter response map for 'Mondrian'. (b) Final filter response map at the end of the segmentation-diffusion process.

luminance edge information derived from the same Gabor-Wavelet representation, and by modifying the domain of integration in the energy functional. This effort will be reported in another paper.

A similar approach can be taken to the problem of speech segmentation: speech segmentation is presently done with either Hidden Markov models or time-warping. We propose that segmentation of time by a Coupled-String model applied to the power spectrum of speech, with couplings between adjacent values of time and frequency, provides a third approach. The Coupled-String model is amenable to dynamic programming and hence fast, and will be effective for all phonemes without the need to model each phoneme in details.

The model uses neurophysiological components as its processing elements, and can be implemented in a locally connected parallel network. There is a strong possibility that it can be linked to the computational processes in the visual cortex. For instance, the segmentation process is related to boundary perception, while the diffusion process can be linked to texture grouping or diffusion phenomenon in psychology. Our work suggests that when cortical complex cells are coupled in a particular fashion, a successive gradient descent type of algorithms can solve a class of image segmentation problems that are essential to visual perception.

References

1. Blake, A. & Zisserman, A. (1987) Visual Reconstruction. The MIT Press, Cambridge, Massachusetts.

2. Bovik, A.C., Clark, M. & Geisler, W.S. (1990) Multichannel Texture Analysis Using Localized Spatial Filters. *IEEE Transections on Pattern Analysis and Machine Intelligence*, Vol. 12, No. 1, January.

3. Brodatz, P. (1966) Texture - A Photographic Album for Artists and Designers. New York: Dover.

4. Daugman, J.G. (1985) Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *J. Opt. soc. Amer.*, Vol. 2, No. 7, pp 1160-1169.

5. Fogel, I. & Sagi, D. (1989) Gabor Filters as Texture Discriminator. *Biological Cybernatics* 61, 103-113.

6. Geiger, D. & Yuille, A. (1991) A Common Framework for Image Segmentation. *Intl. Journal of Computer Vision*, 6:3,227-243.

7. Geman, D., Geman, S., Graffigne, C., & Dong, P. (1990) Boundary Detection by Constrained Optimization. *IEEE Transections on Pattern Analysis and Machine Intelligence*, Vol 12, No. 7, July.

8. Gibson, J.J. (1979) The Ecological Approach to Visual Perception, Houghton-Mifflin.

9. Heeger, D.J. (1989) Computational Model of Cat Striate Physiology. *MIT Media Laboratory Technical Report 125.* October, 1989.

10. Hubel, D.H., & Wiesel, T.N. (1977) Functional Architecture of macaque monkey visual cortex. *Proc. R. Soc. Lond.* B. 198.

11. Kirkpatrick, S., Gelatt, C.D. & Vecchi, M.P. (1983) Optimization by simulated annealing. *Science,220.* 671-680

12. Lee, T.S., Mumford, D. & Yuille, A. (1991) Texture Segmentation by Minimizing Vector-Valued Energy Functionals: The Couple-Membrane Model. *Harvard Robotics Laboratory Technical Report* no. 91-22.

13. Malik, J. & Perona, P. (1989) A computational model for texture segmentation. *IEEE CVPR Conference Proceedings.*

14. Marroquin, J.L. (1984). Surface Reconstruction Preserving Discontinuities (Artificial Intelligence Lab. Memo 792). MIT, Cambridge, MA. A more refined version " A Probabilistic Approach to Computational Vision" appeared in *Image Understanding 1989*. Ed. by Ullman, S. & Whitman Richards. Ablex Publishing Corporation, New Jersey 1990.

15. Mumford, D. & Shah, J.(1985) Boundary Detection by Minimizing Functionals, I. *IEEE CVPR Conference Proceedings* June 19-23. A more detailed and refined version appeared in *Image Understanding 1989*. Ed. by Ullman,S. & Whitman Richards. Ablex Publishing Corporation, New Jersey 1990.

16. Pollen, D.A. & Gaska, J.P. & Jacobson, L.D. (1989) Physiological Constraints on Models of Visual Cortical Function. in *Models of Brain Function*, Ed. Rodney, M.J. Cotterill. Cambridge University Press, England.

17. Reed, T.R. & Wechsler, H. (1990) Segmentation of Textured Images and Gestalt Organization Using Spatial/Spatial-Frequency Representation. *IEEE Transections on Pattern Analysis and Machine Intelligence*, Vol. 12, No. 1, January.

18. Silverman, M.S., Grosof, D.H., De Valois, R.L., & Elfar, S.D. (1989) Spatial-frequency Organization in Primate Striate Cortex. *Proc. Natl. Acad. Sci. U.S.A.*, Vol. 86, January.

19. Voorhees, H. & T. Poggio. (1988) Computing texture boundaries in images. *Nature, 333*:364-367. A detailed version exists as Voorhees' master thesis, "Finding Texture Boundaries in Images". MIT AI Lab Technical Report No. 968, June 1987.

20. Webster, M.A. & De Valois, R.L. (1985) Relationship between Spatial-frequency and Orientation tuning of Striate-Cortex Cells. *J. Opt. Soc. Am.* A Vol 2, No. 7 July 1985.