A Spectral Histogram Model for Textons and Texture Discrimination

Xiuwen Liu
Department of Computer Science
P.O. Box 4530
Florida State University
Tallahassee, FL 32306-4530
liux@cs.fsu.edu

DeLiang Wang
Department of Computer & Information Science
Center for Cognitive Science
The Ohio State University
Columbus, OH 43210-1277
dwang@cis.ohio-state.edu

Abstract

Based on a local spatial/frequency representation, the spectral histogram of an image is defined as the marginal distribution of responses from a bank of filters. We propose the spectral histogram as a quantitative definition for textons. The spectral histogram model avoids rectification and spatial pooling, two commonly assumed stages in texture discrimination models. By matching spectral histograms, an arbitrary image can be transformed via statistical sampling to an image with similar textons to the observed. Texture synthesis is employed to verify the adequacy of the model. Building on the texton definition, we use the χ^2-statistic to measure the difference between two spectral histograms, which leads to a texture discrimination model. The performance of the model well matches psychophysical results on a systematic set of texture discrimination data. A quantitative comparison with the Malik-Perona model is given, and the biological plausibility of the model is discussed.

1 Introduction

Texture perception is one of the pillars in the study of early visual perception. In a life-long effort to pursue a scientific theory for texture perception, Julesz and his colleagues are the most influential in conceptual thinking about texture perception as well as in setting the empirical agenda on the investigation of texture discrimination. After extensive formulations and reformulations in terms of high-order statistics, Julesz eventually proposed the texton theory for texture perception. According to the texton theory, textures are discriminated if they differ in the density of certain simple, local textural features, or textons[11]. Three textons have been consistently specified[10]: elongated blobs defined by color, orientation, size, etc., line terminators, and line crossings. Collinearity and local closure are often mentioned in the literature as well. Though theorized by Julesz as perceptual atoms, "What textons really are is hard to define." (Julesz[11], p.134).

More recently, texture discrimination has received considerable attention from the computational perspective, and many models have been proposed Although these models differ in details, they all share a three-stage common structure[1], referred to as the back-pocket model by Chubb et al[4]. The first stage consists of a set of linear filters, followed by some nonlinear rectification process to remove negative filter responses and compress the range of responses. Note that the rectifying nonlinearity is necessary, since, otherwise, linear filters will give responses that cannot discriminate in the second stage two texture patches with the same mean luminance. The second stage is a pooling stage that performs some spatial averaging, smoothing, or nonlinear inhibition to remove inhomogeneity in rectified filter responses. This stage is necessary because filters are regularly laid out, whereas texture elements on an image are not. Without the second stage filter responses within a homogeneous texture region would produce inhomogeneous responses due to the misalignment between texture elements and filters, creating problems for subsequent processing. The third stage determines texture boundary, or equivalently, produces texture regions on the basis of some edge/contour detection.

The above computational models are mainly motivated by perceptual data obtained from synthetic textures. Though some of these models are quite successful in accounting for empirical data, they do not provide an explicit model for texton or texture. Thus, these quantitative efforts do not seem to provide much insight into the nature of texture perception in Julesz's sense.

In this paper, we propose a model for textons, called

0-7803-7044-9/01/$10.00 ©2001 IEEE

the spectral histogram model. Our model consists of marginal distributions of responses from a bank of filters within an image window. We show that this model elegantly avoids both the rectifying nonlinearity and the pooling stage, thus resulting in a significantly more parsimonious model. The adequacy of the model is established by extensive results on synthesizing both synthetic and natural textures using an effective sampling algorithm. To address texture discrimination, we employ the χ^2-statistic to measure the distance between two spectral histograms. This model yields surprisingly good performance on a systematic set of texture discrimination data. This performance is compared with that of the Malik and Perona model[15]. The spectral histogram model demonstrates the nonlinearity of human texture discrimination. Furthermore, we illustrate that it can exhibit the asymmetry phenomenon in texture discrimination.

This paper is organized as follows. Section 2 describes the spectral histogram model and some of its properties. Section 3 presents our spectral histogram model for textons. Section 4 simulates a set of psychophysical data on texture discrimination, and draws a comparison with the Malik-Perona model. Section 5 discusses the biological plausibility of our model. A technical report with implementation details is also available [14].

2 Spectral Histogram Model

Based on psychophysical and neurophysiological data, it is widely accepted that the human visual system transforms a retinal image into a local spatial/frequency representation. For texture modeling, filter responses themselves are not adequate as textures are regional properties, which is supported by a recent comprehensive study on filter-based methods for texture classification[16]. However, the human visual system perceives various textures without difficulty. Within the spatial/frequency representation, additional steps seem necessary in order to address the inadequacy of filter responses. One reasonable step would be to integrate information from filter responses so as to form perceptually meaningful feature statistics for textures. Studies of human texture perception[2, 4] show that two textures are often perceptually similar when they give a similar distribution of responses from a bank of filters. Recently Heeger and Bergen[9] proposed a texture synthesis algorithm by matching independently the histograms of observed and synthesized image pyramids, which motivated extensive research in texture modeling and synthesis (see e.g., Zhu et al. [21]).

Within the local spatial/frequency representation framework, we define a *spectral histogram* model for characterizing textons, and apply the model to texture discrimination.

Given an input image window $\mathbf{W}$ and a bank of filters $\{F^{(\alpha)}, \alpha = 1, 2, \ldots, K\}$, we compute, for each filter $F^{(\alpha)}$, a sub-band image $\mathbf{W}^{(\alpha)}$ through linear convolution, i.e., $\mathbf{W}^{(\alpha)}(v) = F^{(\alpha)} * \mathbf{W}(v) = \sum_u F^{(\alpha)}(u)\mathbf{W}(v-u)$, whereby a circular boundary condition is used for convenience. For $\mathbf{W}^{(\alpha)}$, we define the marginal distribution, or histogram

$$H^{(\alpha)}_{\mathbf{W}}(z) = 1/|\mathbf{W}| \sum_v \delta(z - \mathbf{W}^{(\alpha)}(v)). \tag{1}$$

We then define the spectral histogram with respect to the chosen filters as

$$H_{\mathbf{W}} = (H^{(1)}_{\mathbf{W}}, H^{(2)}_{\mathbf{W}}, \ldots, H^{(K)}_{\mathbf{W}}).$$

The spectral histogram of an image or an image patch is essentially a vector consisting of marginal distributions of filter responses. The size of the input image window, $|\mathbf{W}|$, is called the *integration scale*. Because the marginal distribution of each filter response is a probability distribution, we define a similarity measure as χ^2-statistic, which is used widely to compare two histograms,

$$\chi^2(H_{\mathbf{W}_1}, H_{\mathbf{W}_2}) = \frac{1}{K} \sum_{\alpha=1}^{K} \sum_z \frac{(H^{(\alpha)}_{\mathbf{W}_1}(z) - H^{(\alpha)}_{\mathbf{W}_2}(z))^2}{H^{(\alpha)}_{\mathbf{W}_1}(z) + H^{(\alpha)}_{\mathbf{W}_2}(z)}. \tag{2}$$

The spectral histogram integrates responses from different filters and provides a naturally normalized feature statistic to compare images of different sizes. By implicitly integrating geometrical and photometric structures of textures, the spectral histogram provides a sufficient model for characterizing perceptual appearance of textures.

3 A Quantitative Definition for Textons

We propose the spectral histogram as a quantitative definition for a texton. The computation leading to a spectral histogram involves commonly used spatial/frequency filters, and thus our definition does not invoke perceptual attributes. Our definition is primarily based on the observation that texture images with a similar histogram are composed of similar elements and similar densities; as such, they would appear perceptually similar, as shown below.

To demonstrate the sufficiency of the spectral histogram, we use 34 filters, including one intensity filter, four local difference filters (two along a row and two along a column), five Laplacian of Gaussian (LoG) filters (with scales of $\sqrt{2}/2$, 1, 2, 4, and 6), and twenty-four Gabor filters (with scales of 2, 4, 6, and 8 and six orientations at each scale), to characterize textons[13]. Given an observed texture, such as the one shown in Fig. 1a, we compute its spectral histogram, which encodes the perceptual structure of the image implicitly. To verify the sufficiency of the spectral histogram for characterizing textures, we generate images by satisfying the constraints $H_{\mathbf{I}} = H_{\text{obs}}$, where $\mathbf{I}$ is an image, $H_{\mathbf{I}}$ its spectral histogram, and H_{obs} the spectral histogram of the observed image. Due to the high dimensionality of $\mathbf{I}$ (for a 128×128 image, the dimension is 16384), the constraints need to be satisfied through stochastic simulation because traditional search methods are computationally not feasible. One commonly used method is the Gibbs sampler[7], which has been demonstrated to be effective for natural textures[20]. Unfortunately, this sampler can be easily trapped at local minima; Fig. 1b shows a typical example of such failure. To explore the image space more effectively, we utilize a sampling procedure from Zhu et al[21]. Here we employ the procedure to generate typical images that share the spectral histogram. Fig. 1c shows the initial condition for the sampling procedure, which is a white noise image. Fig. 1d shows an intermediate image at sweep 500 and Fig. 1e and 1f show synthesized images at sweep 1500 and 4000 respectively, which are perceptually similar to the observed image. The texture element is synthesized very well through the constraints imposed by different filters; the global structure is also reproduced. We stress that while Fig. 1e and 1f are perceptually similar, the corresponding images are quite different in that the positions of crosses vary significantly. Fig. 1g shows the average histogram error per filter with respect to the number of sweeps. As is evident from Fig. 1g, there are several local minimum states, and our sampling procedure overcomes local minima and reaches a globally meaningful state. Fig. 2 shows several more examples, where synthesized texture captures both micropatterns and their spatial layout. These examples demonstrate clearly that our model is sufficient for a variety of textons. Fig. 2f shows an interesting case, where one spectral histogram captures both circle and cross elements at the same time (recall that boundary wrap-around is employed).

Note that the filters are fixed for all the synthesis examples and there is no explicit template for textons. The basic elements are captured by the spectral histograms through imposed constraints by different filters. This offers distinct advantages over texton models based on explicit templates. Not only must a large number of templates be specified to model different kinds of textures, but also must the elements appeared in the observed image be extracted, which is computationally expensive. In addition, a perceptual distance between textures still needs to be defined as textures consisting of different templates need to be compared for discrimination (see Fig. 3 for example).

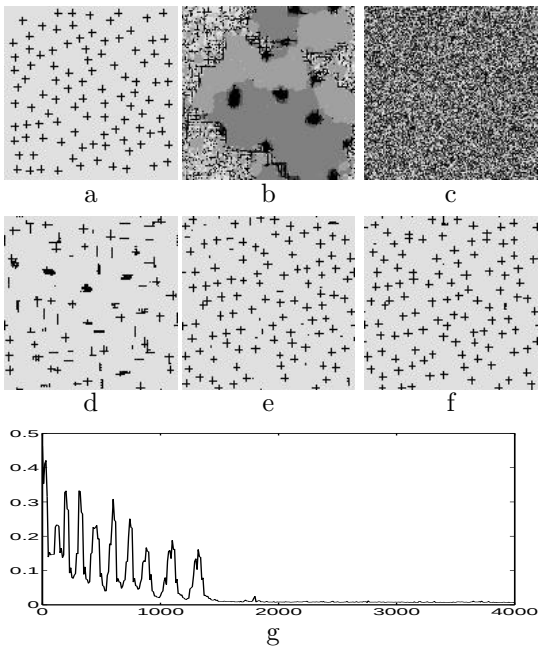

Figure 1: A texture and synthesized images at different sweeps. The size of the image is 128×128. a, Observed image. b, A typical synthesized image using the Gibbs sampler. c, Initial image for sampling, which is white noise. d-f, Synthesized images at sweep 500, 1500, and 4000 respectively. g, The average spectral histogram error per filter with respect to the number of sweeps for the synthesized images.

The above results clearly demonstrate that different images with similar spectral histograms yield perceptually similar appearances. These results on synthetic images, together with extensive results on natural textures[9, 20, 13], suggest that spectral histograms capture a level of image description that is sensitive to certain types of spatial information such as orientation, scale, and density, while oblivious to elaborate geometrical properties. A texture model requires a balance between descriptions that are too simple to reveal anything different and those that are too complex to generate any abstraction of an image. The spectral histogram model, we believe, strikes a balance of this kind.

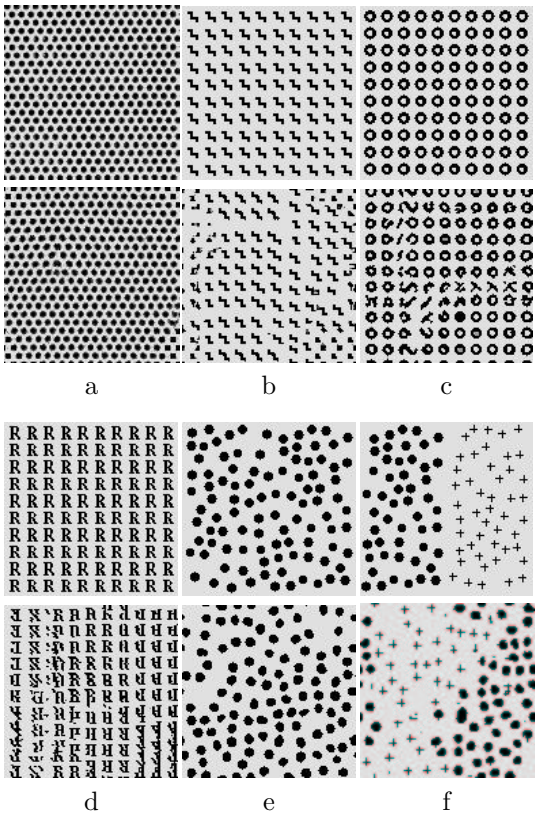

Figure 2: Synthesized images for synthetic textures with different micropatterns. In each column, the upper part shows the observed texture and the lower part a synthesized texture at sweep 2000 from an initial white noise image. The size of all images is 128×128.

4 Texture discrimination

The previous section demonstrates that the spectral histogram model provides a viable definition for textons. Given that much of psychophysical data on texture perception is on comparing texture images, a critical evaluation of any attempt to give a quantitative definition of textons is to match psychophysical data of texture discrimination. This section tests our model with a set of systematic human data on texture discrimination. The set consists of 10 texture pairs, as shown in Fig. 3. The same ten texture pairs were used to evaluate the well-known Malik and Perona model, thus facilitating a quantitative comparison with their model.

We adopt similar procedures used by Malik and Perona [15] for testing texture discrimination. Instead of using 96 filters, we use two gradient filters and three LoG filters, resulting in a total of five filters. At each pixel, we extract local spectral histograms at integration scale 29×29 and the gradient is the average χ^2-

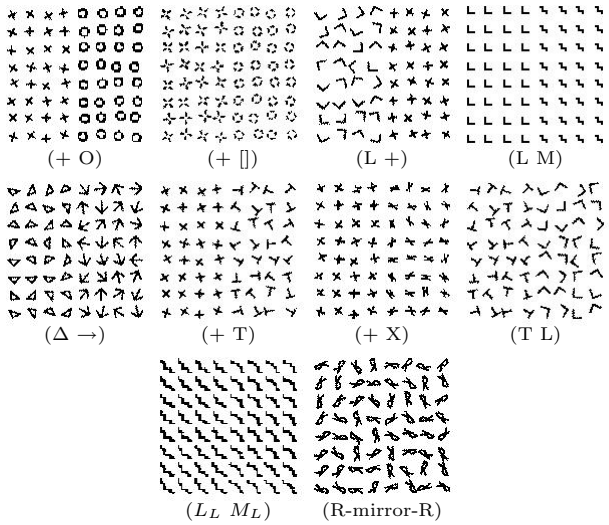

Figure 3: Ten texture pairs scanned from Malik and Perona[15]. The size of all the scanned images is 154×154.

distance per filter between the spectral histograms of the two adjacent windows along a row. Then the gradient is averaged along each column as done in Malik and Perona[15]. The texture gradients generated from our method for the two texture pairs (+ O) and (R-mirror-R) are shown in Fig. 4.

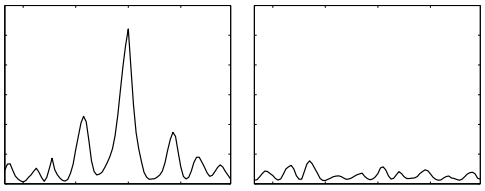

Figure 4: The averaged texture gradient for two selected texture pairs in Fig. 3. Left is the texture gradient averaged along each column for texture pair (+ O) and the right the texture gradient for texture pair (R-mirror-R). The horizontal axis is the column number and the vertical axis is the gradient.

Several observations can be made from the gradient results of Fig. 4. First, a texture pattern does not give rise to a homogeneous texture region, and variations within each texture region are clearly visible. Second, because of the variations among different micropatterns, the absolute value of texture gradient should not be used directly as a measure for texture discrimination as in Malik and Perona [15]. Based on these observations, we propose a texture discrimination measure as the difference between the central peak and the maximum of two adjacent side peaks. In other words, the

peak corresponding to the middle boundary is compared against the two adjacent ones corresponding to the interior boundaries within each texture. In the (+ O) case, the central peak is 0.239, and the left and right side peaks are 0.104 and 0.08 respectively. Thus the discrimination measure is 0.135. For the (R-mirror-R) case, the central peak is 0.017 and the left and right side peaks are 0.012 and 0.027 respectively. Thus the measure is -0.01, indicating that the two texture regions are not discriminable at all. We calculate the proposed discrimination measure for the ten texture pairs. Table 1 shows our results along with the psychophysical data from Kröse[12], and the results from Malik and Perona's model [15]. Here the data from Kröse[12] was based on the converted data given in Malik and Perona[15]. Fig. 5 shows the data points linearly scaled so that the measures for the second pair (+ []) match. Our measure predicts that that (+ O) is much easier to discriminate than all the other pairs, the pair (L_L, M_L) is barely discriminable with a score of 0.001, and the pair (R-mirror-R) is not discriminable with a score of -0.01.

Table 1: Texture Discrimination Scores

Texture pair	Texture discriminability		
	Human Data [12]	Malik and Perona [15]	Our Results
(+ O)	100	407	0.135
(+ [])	88.1	225	0.036
(L +)	68.6	203	0.027
(L M)	n.a.	165	0.023
(Δ $\rightarrow$)	52.3	159	0.018
(+ T)	37.6	120	0.015
(+ X)	30.3	104	0.014
(T L)	30.6	90	0.004
(L_L, M_L)	n.a.	85	0.001
(R-mirror-R)	n.a.	50	-0.01

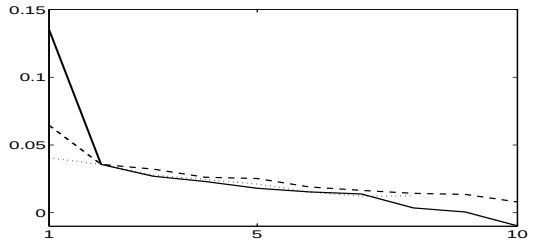

Figure 5: Texture discrimination results. Here the horizontal axis corresponds to the order of the texture pairs in Table 1 and the vertical axis the texture discrimination scores. Dotted line - Psychophysical data from Krose[12]; dash line - results from Malik and Perona's model[15]; solid line - results from the spectral histogram model.

It is clear from Table 1 that our model performance is entirely consistent with the other two. We employ only 5 commonly used filters instead of 96 filters in their model. The Malik and Perona model needs an elaborate form of nonlinearity that depends on inter-filter interactions specific to different filter types.

Nonlinearity is an important property of human texture discrimination. Texture pairs (L, M) and (L_L, M_L) were constructed by Williams and Julesz [18] to argue against linear models. The (L, M) pair is among the ones that are easily discriminable. However, the (L_L, M_L) pair, which was constructed by adding a uniform texture of little L's at the endpoints of the L's and M's in the (L, M) pair, is not. This demonstrates that the texture discrimination cannot be a simple linear operation; some form of nonlinearity must be included in order to account for this phenomenon. The Malik and Perona model[15] is able to reproduce this nonlinearity by incorporating two nonlinear stages. In contrast, our model reproduces the nonlinearity without any additional nonlinear operation.

According to Malik and Perona[15], their model cannot account for asymmetry in texture discrimination, which refers to the phenomenon that one texture embedded in another one is more readily discriminated than if the latter is embedded in the former[19]. Our model may be able to account for the asymmetry phenomenon based on the following observation. The variability within one texture is generally different from the variability within the other when asymmetry is present and our discrimination score depends on the variability. A preliminary experiment supports this conclusion [14].

5 Biological Plausibility

The filtering stage in the spectral histogram model is commonly used in other models of texture discrimination, and as mentioned earlier the existence of early processing by spatial/frequency filters in the visual system is widely accepted[5].

To compute a histogram requires neurons with sizable receptive fields, which would presumably occur in the extrastriate cortex[17]. Such cells would summate similar responses from a pool of neurons behaving as filters, which would presumably be within V1. Neurobiological findings show that neurons in V1 are interconnected by long-range horizontal connections, and horizontal connections link neurons with similar orientation attributes [8]. Horizontal connections of this kind facilitate histogram gathering. The spectral histogram model further suggests that responses of all magnitudes within

a pool of similar filters be collected, not just optimal responses.

The above interpretation relies on some spatial summation. Another possible interpretation invokes temporal coding, which would require more elaborate timing but be more parsimonious. The idea is that spectral histograms, are entirely coded by temporal responses of single neurons. This would require a neuronal mechanism that responds to different filters with different latencies so that the responses from distinct filters would not be confused. Experimental studies have suggested that neuronal latencies may encode perceptual information[6]. Also, evidence from visual perception and neurobiology suggests systematic and concerted temporal coding[3].

Given coded histograms, a distance measure between them such as the χ^2-statistic could be readily performed by a correlation mechanism. Neurons performing a correlational analysis have been implicated in many contexts.

Acknowledgments This research was supported in part by an ONR Young Investigator Award (N00014-96-1-0676), an NSF grant (IIS-0081058), and an AFOSR grant (F49620-01-1-0027) to DLW.

References

[1] J. Beck, "Textural segmentation," in *Organization and Representation in Perception*, J. Beck, ed., Erlbaum, Hillsdale, NJ, pp. 285–318, 1982.

[2] J. R. Bergen and E.H. Adelson, "Early vision and texture perception," *Nature*, vol. 333, pp. 363–367, 1988.

[3] G. Buzsáki, R. Llinás, W. Singer, A. Berthoz, and Y. Christen, *Temporal Coding in the Brain*, Springer-Verlag, Berlin, 1994.

[4] C. Chubb, J. Econopouly, and M. S. Landy, "Histogram contrast analysis and the visual segregation of IID textures," *J. Opt. Soc. Am. A*, vol. 11, pp. 2350–2374, 1994.

[5] R. L. De Valois and K. K. De Valois, *Spatial Vision*, Oxford University Press, New York, 1988.

[6] T. J. Gawne, T. W. Kjaer, and B. J. Richmond, "Latency: Another potential code for feature binding in striate cortex," *J. Neurophysiol.*, vol. 76, pp. 1356–1360, 1996.

[7] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 6, pp. 721–741, 1984.

[8] C. D. Gilbert and T. N. Wiesel, "Columnar specificity of intrinsic horizontal and corticocortical connections in cat visual cortex," *J. Neurosci.*, vol. 9, pp. 2432–2442, 1989.

[9] D. J. Heeger and J. R. Bergen, "Pyramid-based texture analysis/synthesis," in *Proceedings of SIGGRAPHS*, pp. 229–238, 1995.

[10] B. Julesz, "Texton gradients: The texton theory revisited," *Biol. Cybern.*, vol. 54, pp. 245–251, 1986.

[11] B. Julesz, *Dialogues on Perception*, MIT Press, Cambridge, MA, 1995.

[12] B. J. Kröse, "A description of visual structure," Ph.D. Dissertation, Delft University of Technology, The Netherlands, 1986.

[13] X. Liu, "Computational investigation of feature extraction and image organization," Ph.D. Dissertation, Ohio State University, Columbus, OH, 1999 (available online at ftp://ftp.cis.ohio-state.edu under pub/tech-report/2000/TR18.pdf.gz)

[14] X. Liu and D. L. Wang, "A spectral histogram models for textons and texture discrimination," Technical Report OSU-CISRC-4/01-TR05, Ohio State University, Columbus, OH, 2001 (available online at ftp://ftp.cis.ohio-state.edu/pub/tech-report/2001/TR05.pdf.gz).

[15] J. Malik and P. Perona, "Preattentive texture discrimination with early vision mechanisms," *J. Opt. Soc. Am. A*, vol. 7, pp. 923–932, 1990.

[16] T. Randen and J. H. Husoy, "Filtering for texture classification: A comparative study," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 21, pp. 291–310, 1999.

[17] P. H. Schiller, "Visual processing in the primate extrastriate cortex," in *Early Vision and Beyond*, T. V. Papathomas, et al., ed., MIT Press, Cambridge, MA, pp. 167–176, 1995.

[18] D. Williams and B. Julesz, "Filters versus textons in human and machine texture discrimination," in *Human and Machine Perception*, H. Wechsler, ed., Academic Press, San Diego, CA, pp. 145–175, 1992.

[19] D. Williams and B. Julesz, "Perceptual asymmetry in texture perception," *Proc. Natl. Acad. Sci. USA*, vol. 89, pp. 6531–6534, 1992.

[20] S. C. Zhu, X. Liu, and Y. N. Wu, "Exploring texture ensembles by efficient Markov Chain Monte Carlo," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 22, pp. 554-569, 2000.

[21] S. C. Zhu, Y. N. Wu, and D. Mumford, "Minimax entropy principle and its application to texture modeling," *Neural Comp.*, vol. 9, pp. 1627–1660, 1997.