Technical Report OSU-CISRC-4/01-TR05
Department of Computer and Information Science
The Ohio State University
Columbus, OH 43210-1277

ftp site: **ftp.cis.ohio-state.edu**
Login: **anonymous**
Directory: **pub/tech-report/2001**
File in pdf format: **TR05.pdf**
web site: **http://www.cis.ohio-state.edu/Research/tech-report.html**

A Spectral Histogram Model for Textons and Texture Discrimination

Xiuwen Liu
Department of Computer Science
Florida State University
Tallahassee, FL 32306-4530, USA
Tel: (850) 644-0050 Fax: (850) 644-0058
Email: liux@cs.fsu.edu
http://www.cs.fsu.edu/∼liux

DeLiang Wang
Department of Computer and Information Science and
Center for Cognitive Science
The Ohio State University
2015 Neil Avenue, 395 Dreese Lab.
Columbus, OH 43210-1277, USA
Tel: (614) 292-7402 Fax: (614) 292-2911
Email: dwang@cis.ohio-state.edu
http://www.cis.ohio-state.edu/∼dwang

A Spectral Histogram Model for Textons and Texture Discrimination

Xiuwen Liu and DeLiang Wang*

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. Furthermore, the model exhibits the nonlinearity and asymmetry phenomena in human texture discrimination. A quantitative comparison with the Malik-Perona model is given, and a number of issues regarding the model are discussed.

1 Introduction

Texture perception is one of the pillars in the study of early visual perception[24, 2]. Much of the psychophysical work concentrates on texture discrimination, or detecting whether two texture patches can be discerned rapidly by human observers (for reviews see Bergen[4]; Papathomas et al.[35]). Effortless texture discrimination takes place rapidly and is viewed as a preattentive process that occurs in parallel across the whole visual field. A critical empirical issue is what stimulus conditions result in preattentive texture segregation as opposed to a slow, effortful process that requires focal attention. Many texture patterns have been devised to test various ideas and hypotheses on this issue, and have revealed an array of perceptual phenomena concerning texture discrimination.

Beck, a pioneer in texture perception, described a multistage conceptual model for texture segregation in 1982. According to his model[3], the first stage performs local feature detection with receptive fields in the visual system. The second stage extracts the total differences in color, luminance, orientation, and size between neighboring texture elements. The last

*Please address all correspondence to Prof. Deliang Wang, Department of Computer and Information Science, The Ohio State University, Columbus, OH 43210-1277. Email: dwang@cis.ohio-state.edu. Phone: (614) 292-6827. Fax: (614) 292-2911.

stage segregates an image into regions of the same texture on the basis of the magnitude and distribution of difference signals.

In a life-long effort to pursue a scientific theory for texture perception similar to that of the Young-Helmholtz trichromatic theory for color 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[25, 27]. Three textons have been consistently specified[25, 26]: 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[27], p.134). As pointed out by Bergen and Adelson[5], one major difficulty with his approach is that "it is based on a verbal description of image features."

More recently, texture discrimination has received considerable attention from the computational perspective, and many models have been proposed (see among others Caelli[7]; Turner[41]; Voorhees and Poggio[43]; Fogel and Sagi[14]; Malik and Perona[32]; Rubenstein and Sagi[39]; Graham et al.[18]; Räth and Morfill[37]; Barth et al.[1]). Although these models differ in details, they all share a three-stage common structure, referred to as the back-pocket model by Chubb et al[9]. 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. We note that this common structure is very similar to Beck's conceptual model described earlier.

The above computational models are mainly motivated by perceptual data obtained from synthetic textures, which are easy to modify in a controlled way - often necessary for collecting systematic psychophysical data. Moreover, these models mainly aim at simulating perceptual data or exhibiting perceptual phenomena. Though some of these models are quite successful in accounting for empirical data, they do not provide an explicit model for texton or texture. All they do is to produce texture boundaries, and their adequacy as a texture model is never checked. Thus, these quantitative efforts do not seem to provide much insight into the nature of texture perception in Julesz's sense.

On the other hand, explicit texture modeling has been undertaken in computer vision. The most popular is a class of Markov Random Field (MRF) models[11, 16, 13]. MRF models define probability distributions based on some correlation within a local neighborhood. Once an MRF model and its parameters are specified, a statistical sampler can be used to generate

(synthesize) individual textures that realize the model. Texture models of this kind have been successfully applied to texture image processing, including classification and segmentation. To our knowledge, however, no study reports any successful attempt to model human texture discrimination. One reason is that these models cannot account for the kind of textures often used in psychophysical experiments; such textures, though synthetic and often binary, generally have some regular structure, and as a result they are very difficult to synthesize on the basis of local statistical models.

In this paper, we propose a mathematical model for textons, called 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. 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 verifies its adequacy as a texton model by synthesizing texture patterns with distinct spectral histograms. Section 4 simulates a set of psychophysical data on texture discrimination, and draws a comparison with the Malik-Perona model. Section 5 compares our model with other texture discrimination models, points out its relations with other studies, and discusses its biological plausibility. Section 6 concludes the paper.

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. Such a representation can be simulated by a bank of filters with tuned frequencies and orientations, e.g. Gabor filters, and finds applications in many areas including image compression. 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[36]. The result shows that all of the filter-based methods included in the study fail to produce meaningful classification results for a set of textures, suggesting that filter responses are not sufficient to characterize textures.

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[5, 9] 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[21] proposed a texture synthesis algorithm by matching independently

the histograms of observed and synthesized image pyramids, which motivated extensive research in texture synthesis. Within the local spatial/frequency representation framework, we define a *spectral histogram* model for characterizing textons, and apply the model to texture discrimination.

2.1 Definition and Properties

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. Two examples of spectral histograms are shown in Fig. 1. 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 and has been used successfully in modeling natural textures[30].

Formally, the spectral histogram model exhibits desired properties for perceptual modeling of textons. Because filter responses depend only on relative locations of pixels, the exact position of the image window $\mathbf{W}$ does not affect its spectral histogram as long as it encloses the same texture region, thus resulting in translation invariance. Because the histogram operation given in Eq. (1) is nonlinear, the spectral histogram is a nonlinear operator. In addition, multiple filters impose different constraints on the geometrical structures of images within the same spectral histogram, and this makes linear summation an operation not applicable to spectral histograms. The nonlinearity of human texture discrimination was demonstrated by Williams and Julesz[46], and was used as evidence against any linear model[5]. The issue of nonlinearity will be further discussed in Sect. 4.

We have shown elsewhere that the spectral histogram can uniquely represent an image up to a translation given sufficient filters[31]. With appropriately chosen filters, the spectral histogram provides a unified texture feature statistic, where many existing texture features can be treated as special cases. See Liu[30] for filter selection and results on texture classification and segmentation.

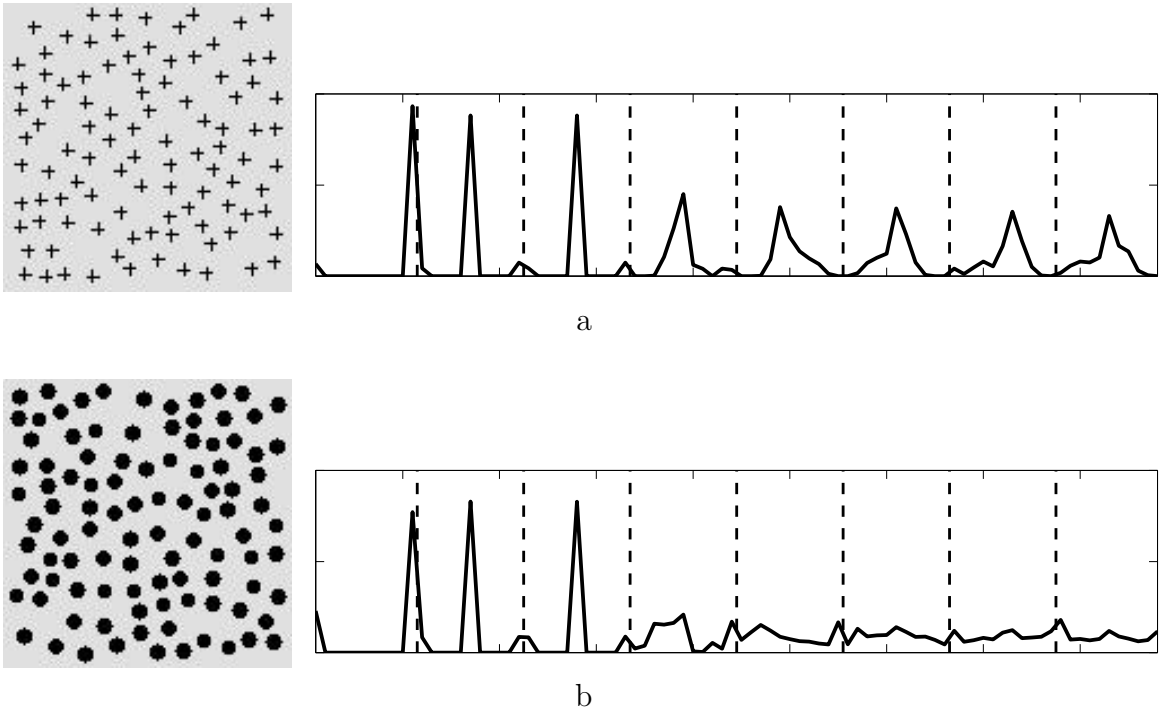
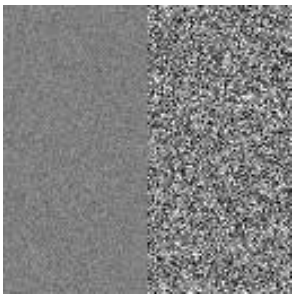


Figure 1: The spectral histograms of two synthetic textures. The size of the images is 128×128. Here eight filters are used to calculate the spectral histogram and they are separated by dash lines. a, A texture consisting of crosses randomly placed. b, A texture consisting of circles randomly placed.

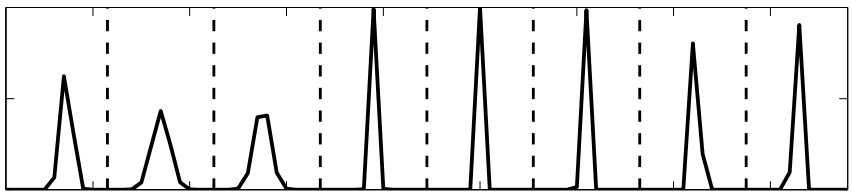
2.2 Relations to Other Texture Discrimination Models

Before demonstrating the performance of the spectral histogram model, we discuss its relations to standard (or back-pocket) models of texture discrimination. Essentially our model consists of two stages: a filtering stage and a histogram gathering and comparison stage. Our first stage is similar to that of the back-pocket model, except that no rectification is done in our model. The reason we do not need a rectifying nonlinearity is that the spectral histograms can reflect the difference in higher-order statistics. To explain this point, we show the histogram responses to two images of identical mean but different variances. Such examples are commonly used to justify the rectifying nonlinearity. Fig. 2 shows an image with the spectral histograms of the left and right half. The image was generated by adding Gaussian noise with different variances to a uniform image and thus the left and right regions have identical mean but different variances. However, their spectral histograms are very different in the relative heights of peaks, resulting in a large χ^2-distance between them. This example illustrates that a histogram can discriminate higher-order statistics, thus no need for rectification. In theory, with enough data and histogram bins, a histogram can provide an approximation of any distribution with arbitrary accuracy[31].

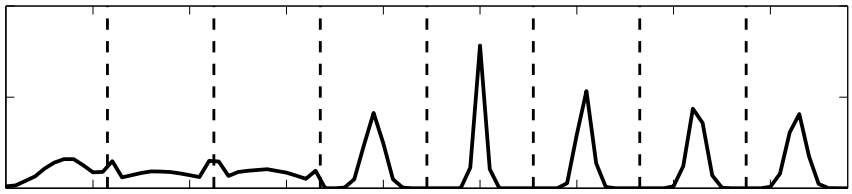
The pooling stage in the back-pocket model is necessary for reducing inhomogeneity in filter responses. Because a histogram is gathered from a window of a particular integration scale, it implicitly performs a kind of spatial smoothing. Since a bin in a histogram counts



a



b



c

Figure 2: The spectral histograms of two regions with an identical mean but different variances. a, An image consisting of such two regions, which is generated by adding Gaussian noise with different variances to a uniform image. The left region has a variance of 10 and the right 50. The size of the image is 128×128. b, The spectral histogram of the left region. c, The spectral histogram of the right region.

how many of the identical filters generate a similar response within a spatial window that is substantially larger than the size of a texture element, this measure is fundamentally insensitive to precise locations of texture elements within the window. This property is illustrated in Fig. 3. Fig. 3a and 3b show two textures with similar spectral histograms, and thus the images would belong to the same texture. Pixelwise, however, the two images are very different. For example, the root-mean-square distance between them is larger than that between Fig. 3a and 3c, the latter being a Gaussian noise image. We emphasize that this important characteristic is consistent with the evidence from human texture discrimination that only the number (or density) of textons has perceptual significance and their positions are ignored (Julesz[25], p. 97).

The stage for computing a texture boundary is similar between our model and the back-pocket model, which involves a difference (or similarity) measure. In our case, because

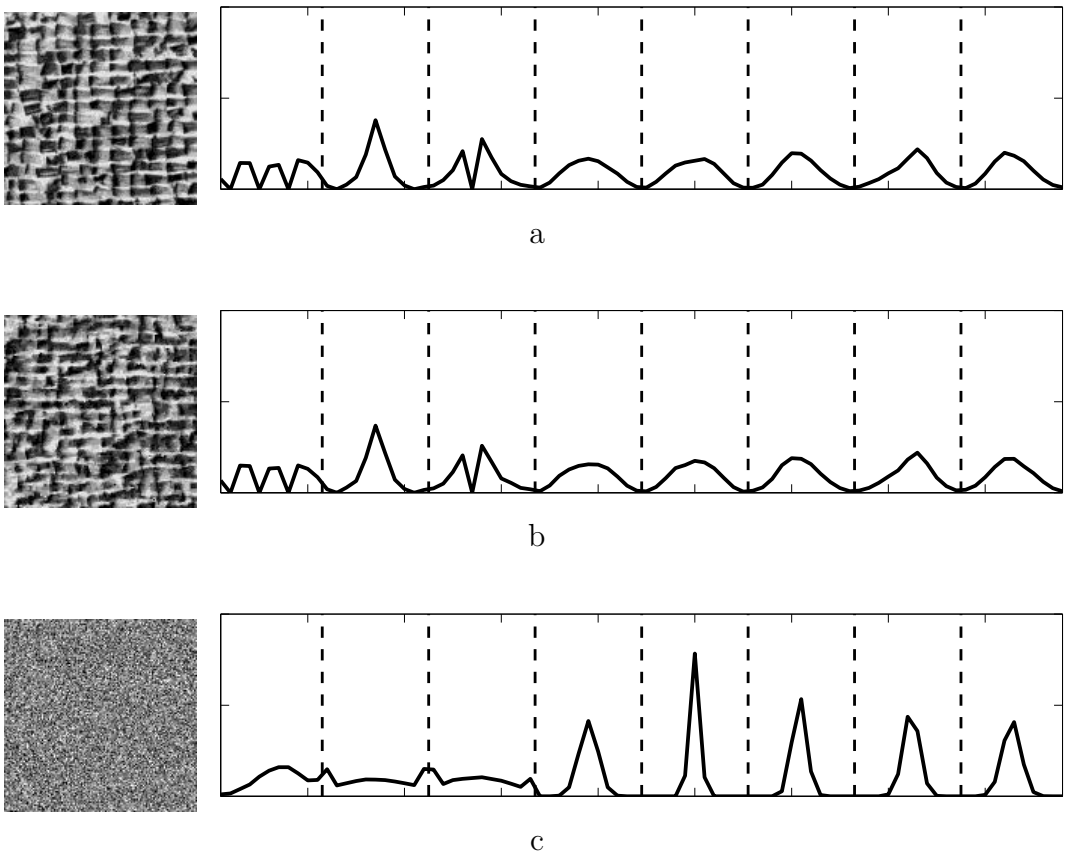


Figure 3: Patches with similar histograms that are perceptually indiscriminable and those with disimilar histograms that are perceptually different. a and b, Two patches with their corresponding spectral histograms. The spectral histograms are similar. However, the root-mean-square distance between the two patches is large - 94.0 per pixel. c, A Gaussian noise image with its spectral histogram. The root-mean-square distance between this patch and that in a is 84.5 per pixel, smaller than the distance between a and b. The size of all the images is 128×128.

histograms are marginal distributions, we use χ^2-statistic to measure the difference between histograms (other statistics can also be used, see Liu[30]). For back-pocket models, the comparison is between two spectra of various filter responses.

The above discussion makes it clear that our model is significantly simpler at the conceptual level. It eliminates a rectifying nonlinearity, and a pooling stage which has been argued to require another nonlinear operation[32]. The computational functions of such operations are intrinsically incorporated in the spectral histogram model. Thus, our model is more parsimonious.

3 Textons as Spectral Histograms

3.1 A Quantitative Definition for Textons

Besides the problem of being a verbal description[5], the notion of textons as conspicuous local features implies that textons are perceptual properties. This seems at odds with the evidence that texture segregation takes place at a level earlier than the one at which perceptual features can be derived[4]. Even for visual cortical cells with Gabor-like receptive fields, which are frequently taken as edge or line detectors, they respond also to sinusoidal gratings, white noise, and many other patterns. Textons such as corner and closure detectors are more specialized and complex to compute, and thus would presumably arise even later in the visual processing pathway.

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[30]. Given an observed texture, such as the one shown in Fig. 4a, 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[16], which has been demonstrated to be effective for natural textures[48]. Unfortunately, this sampler can be easily trapped at local minima; Fig. 4b shows a typical example of such failure. To explore the image space more effectively, we utilize a sampling procedure from Zhu et al[49]. Their procedure was originally proposed to learn parameters in a probability model derived from the maximum entropy principle[22]. By changing the parameters adaptively, the resulting sampling procedure more effectively explores the space in which the spectral histogram matches the observed one. Here we employ the procedure to generate typical images that share the spectral histogram. Fig. 4c shows the initial condition for the sampling procedure, which is a white noise image. Fig. 4d shows an intermediate image at sweep 500 and Fig. 4e and 4f 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. 4e and 4f are perceptually similar, the corresponding images are quite different in that the positions of crosses vary significantly. This demonstrates that our sampling algorithm moves among images that tend to be similar perceptually, but not literally (see also Fig. 3). Fig. 4g shows

the average histogram error per filter with respect to the number of sweeps. As is evident from Fig. 4g, there are several local minimum states, and our sampling procedure overcomes local minima and reaches a globally meaningful state.

Fig. 5 shows another example. Fig. 5a shows the observed image, which consists of randomly placed circles. Fig. 5b-5d show the synthesized results from three different initial conditions. By matching with the observed spectral histogram, all the three images are transformed into images that are perceptually similar to the observed. The circles are reproduced well, and so is the density of the circles.

Fig. 6 shows several more examples. Fig. 6a shows a synthesized texture consisting of hexagons arranged regularly. The spectral histogram captures the hexagons as well as the regular structure. Fig. 6b shows a texture consisting of a widely used micropattern in texture discrimination. The micropatterns are captured by the spectral histogram. Fig. 6c shows a texture of empty circles placed on a regular grid. The structure of each element is synthetized solely based on the spectral histogram. Fig. 6d shows that 'R' can be reproduced using the spectral histogram. Worth noting in the above examples is that the regular layout of texture elements is very well captured by the spectral histogram model. Fig. 6e demonstrates that the spectral histogram can capture textures with low-density elements. Fig. 6f 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[43]. 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. 8 for example).

Our proposed texton model is perceptually sufficient not only for synthetic texture patterns, but also for natural images, as shown by Heeger and Bergen[21], Liu[30], and Zhu et al[48]. For example, Fig. 7a shows a cheetah image and Fig. 7b shows a patch of cheetah skin. Fig. 7c shows the synthesized patch by matching the spectral histograms. The synthesized image captures the perceptual characteristics of the cheetah skin.

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, 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[45]. The spectral histogram model, we believe, strikes a balance of this kind.

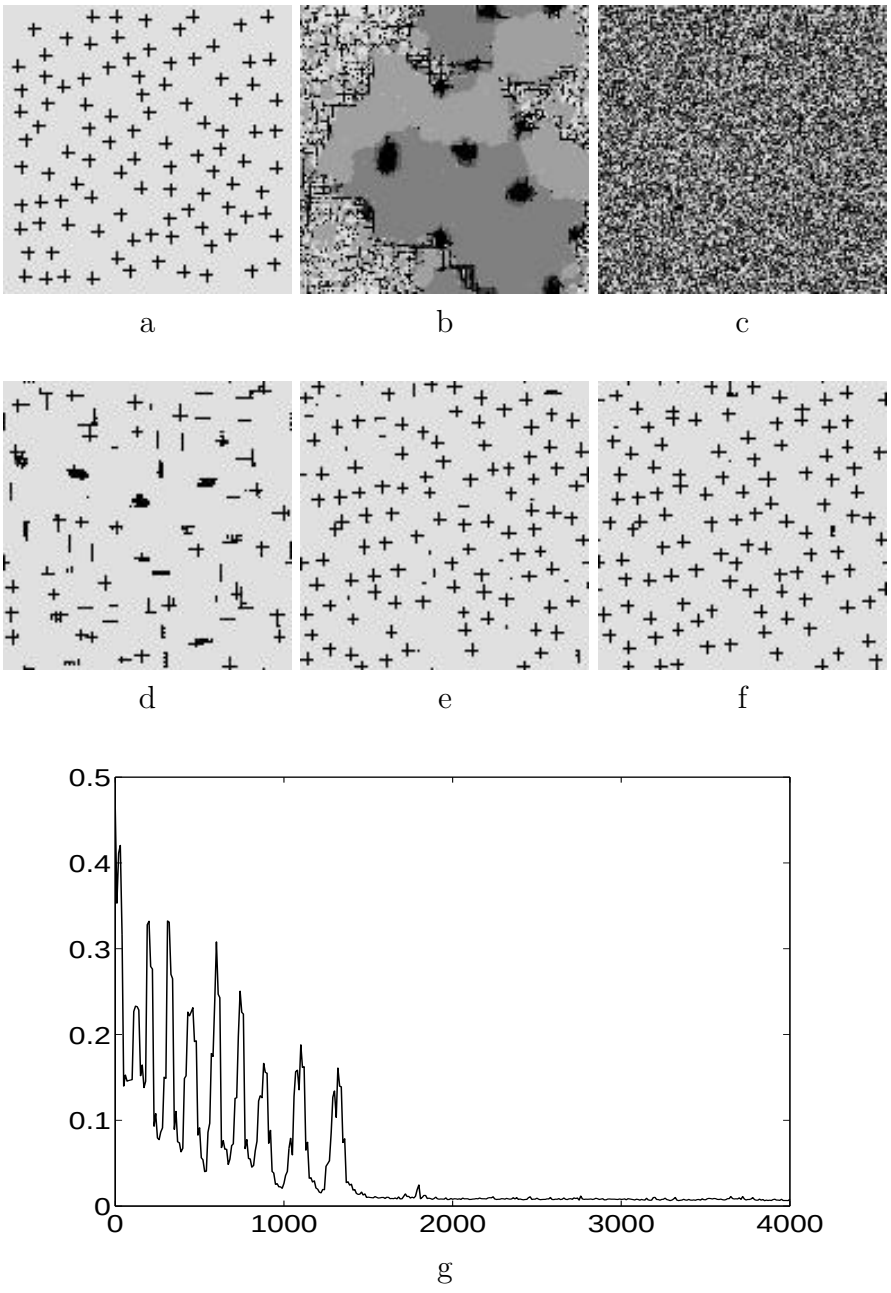


Figure 4: A texture and synthesized images at different sweeps. The size of the image is 128×128. a, Observed image. b, A typical synthesized image using the Gibbs sampler. c, Initial image for sampling, which is white noise. d-f, Synthesized images at sweep 500, 1500, and 4000 respectively. g, The average spectral histogram error per filter with respect to the number of sweeps for the synthesized images.

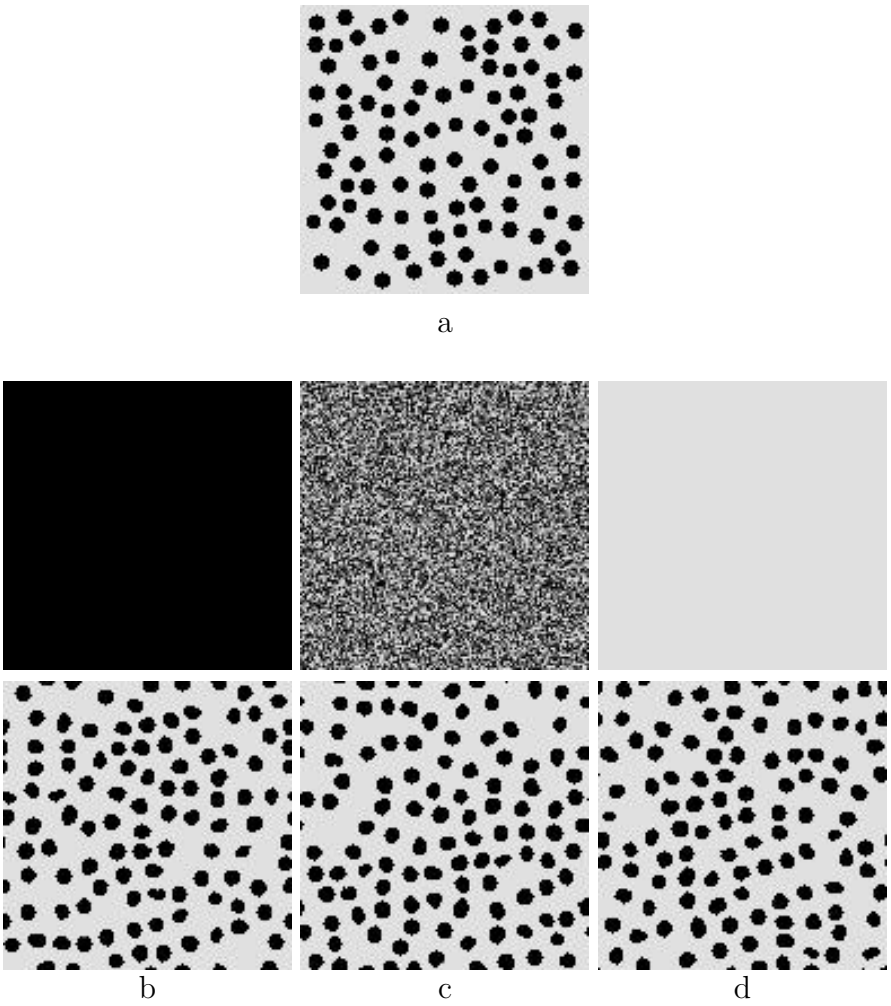


Figure 5: A texture and synthesized images from different initial conditions. The size of the image is 128×128. Note that the circular boundary condition is used when synthesizing images and synthesized circles on the boundaries should be viewed as continuing across the boundaries. a, Observed texture. b-d, The lower part shows the synthesized images at sweep 2000 from the corresponding initial images of the upper part.

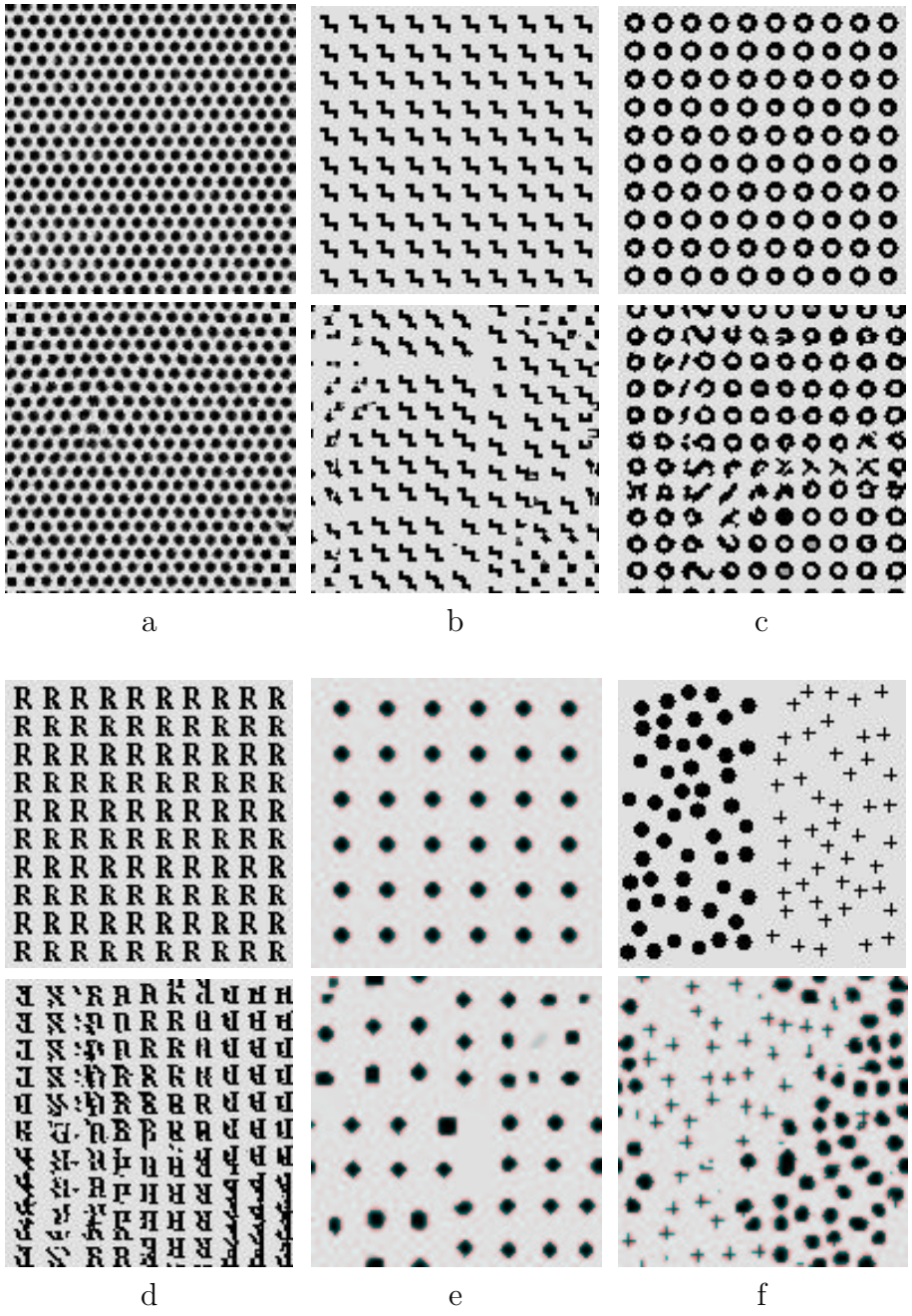


Figure 6: Synthesized images for synthetic textures with different micropatterns. In each column, the upper part shows the observed texture and the lower part a synthesized texture at sweep 2000 from an initial white noise image. The size of all images is 128×128. a, A texture consisting of regularly arranged hexagons. b, A texture consisting of a micropattern widely used in texture discrimination experiments. c, A texture with empty circles. d, A texture consisting of R's. e, A texture consisting of filled circles. f, An image consisting of two distinct textures. The spectral histogram model captures both texture elements and their spatial layout.

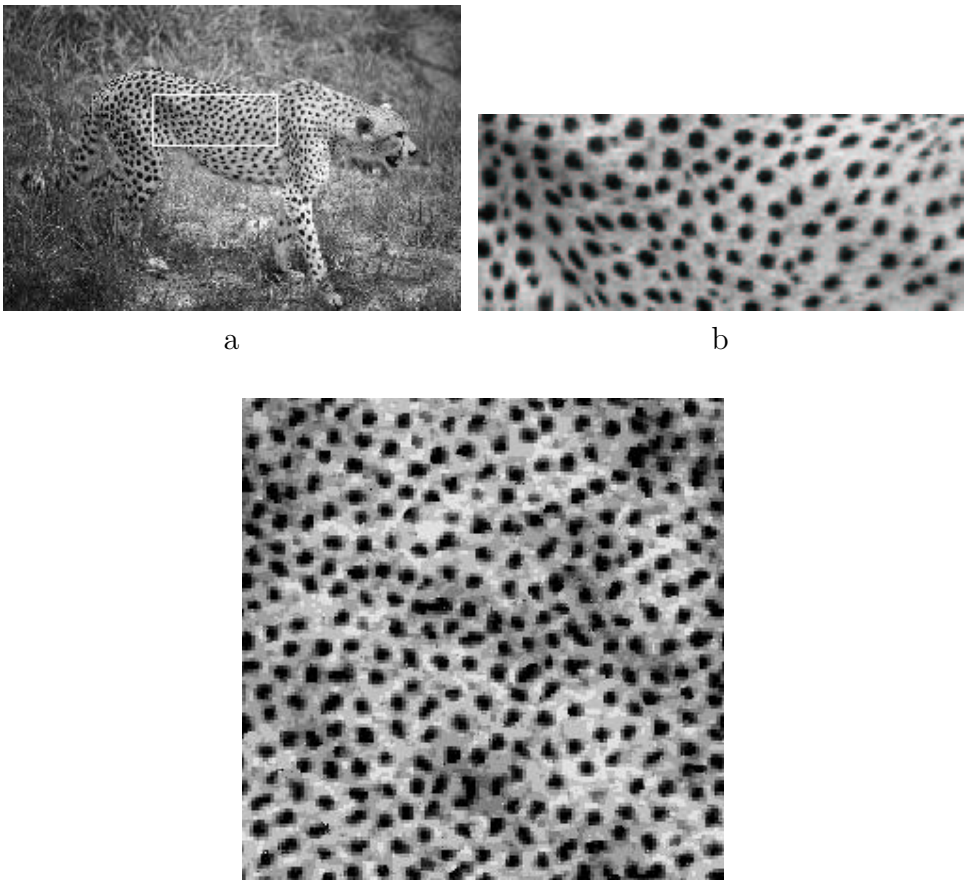


Figure 7: Natural texture of cheetah skin. a, An image containing a cheetah. The size of the image is 648×972. b, The cheetah skin from the enclosed area in a. The size of this area is 104×258. c, A synthesized image, with the size 256×256.

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. 8. Seven are from Kröse[29], two from Williams and Julesz[46], and one composed of R's and mirror-image R's (called R-mirror-R). The same ten texture pairs were used to evaluate the well-known Malik and Perona model, thus facilitating a quantitative comparison with their model. The texture pairs shown in Fig. 8 were scanned from Malik and Perona[32].

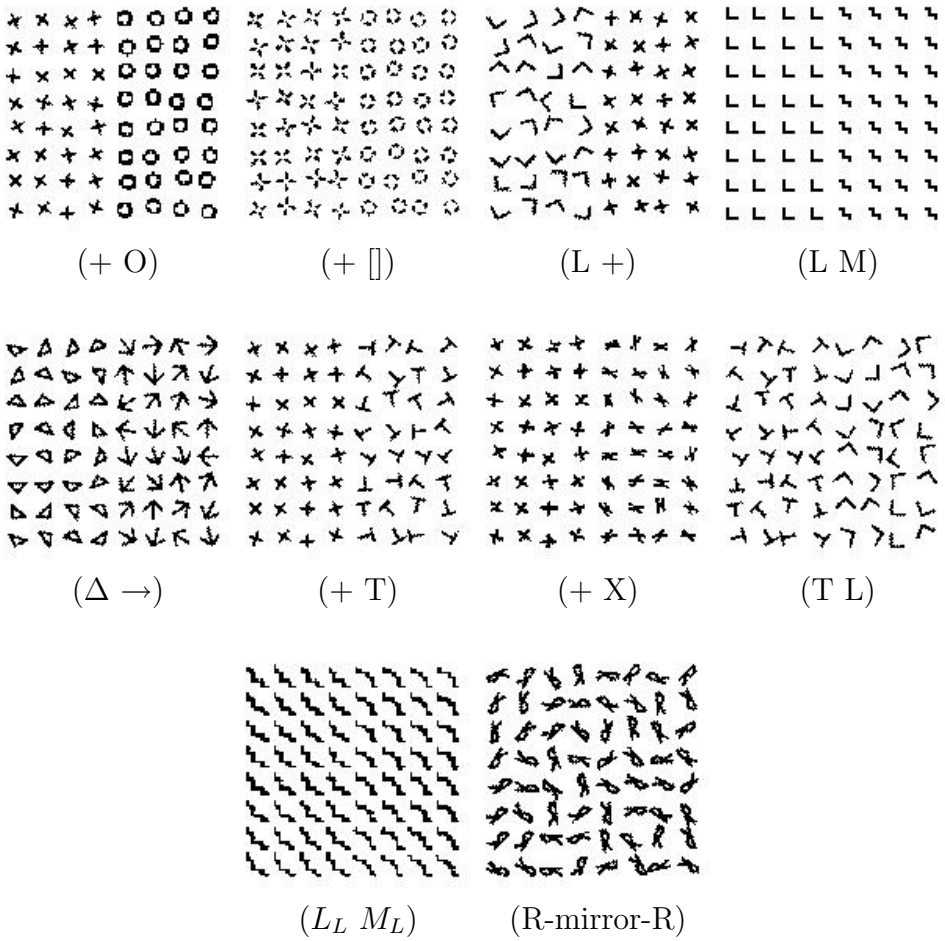


Figure 8: Ten texture pairs scanned from Malik and Perona[32]. The size of all the scanned images is 154×154.

We adopt similar procedures used by Malik and Perona [32] for testing texture discrimination. Instead of using 96 filters, we use two gradient filters (D_{xx} and D_{yy})[30] and three

LoG filters with scales $\sqrt{2}/2$, 1, and 2, 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-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[32]. The texture gradients generated from our method for the two texture pairs (+ O) and (R-mirror-R) are shown in Fig. 9b and 9d.

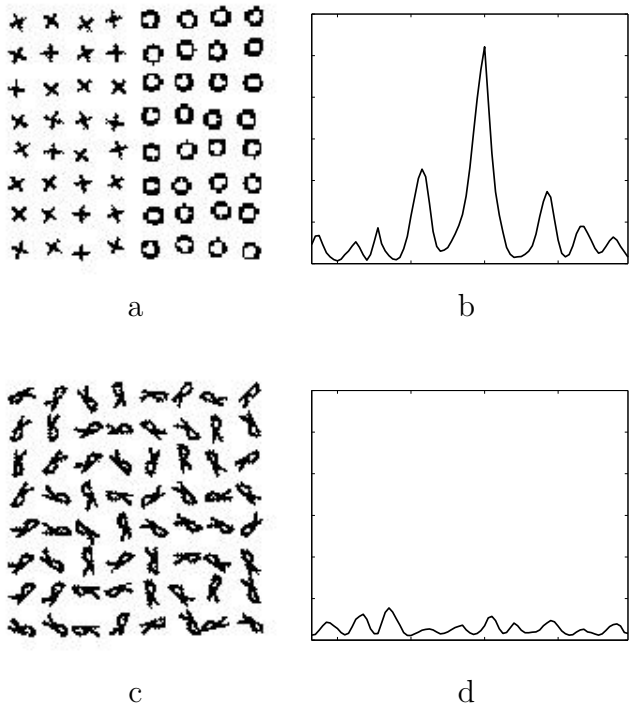


Figure 9: The averaged texture gradient for two selected texture pairs in Fig. 8. a, The texture pair (+ O). b, The texture gradient averaged along each column for a. The horizontal axis is the column number and the vertical axis is the gradient. c, The texture pair (R-mirror-R). d, The texture gradient for c.

Several observations can be made from the gradient results of Fig. 9. First, a texture pattern does not give rise to a homogeneous texture region, and variations within each texture region are clearly visible. For regularly arranged micropatterns people do perceive distinct columns besides the middle boundary that separates two main texture regions; see the texture pair (+ O) in Fig. 8. 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 [32]. As shown in Fig. 9d, even though the gradient is much weaker compared to Fig. 9b, the filters still respond to element variations, which is also evident in Malik and Perona [32]. However, no texture boundary is perceived in this case.

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[29], and the results from Malik and Perona's model [32]. Here the data from Kröse[29] was based on the converted data given in Malik and Perona[32]. Fig. 10 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 [29]	Malik and Perona Results [32]	Spectral Histogram 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

It is clear from Table 1 that our model performance is entirely consistent with the other two. Note that the Malik and Perona model[32] is a representative of back-pocket models; thus our qualitative comparisons in Sect. 2.2 apply to their model. In addition, we employ only 5 commonly used filters instead of 96 filters in their model. Their model needs an elaborate form of nonlinearity that depends on inter-filter interactions specific to different filter types. (They reported that simplified versions of this nonlinearity produce inferior performances.)

As alluded to earlier, nonlinearity is an important property of human texture discrimination. Texture pairs (L, M) and (L_L, M_L) were constructed by Williams and Julesz [46] 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[32] is

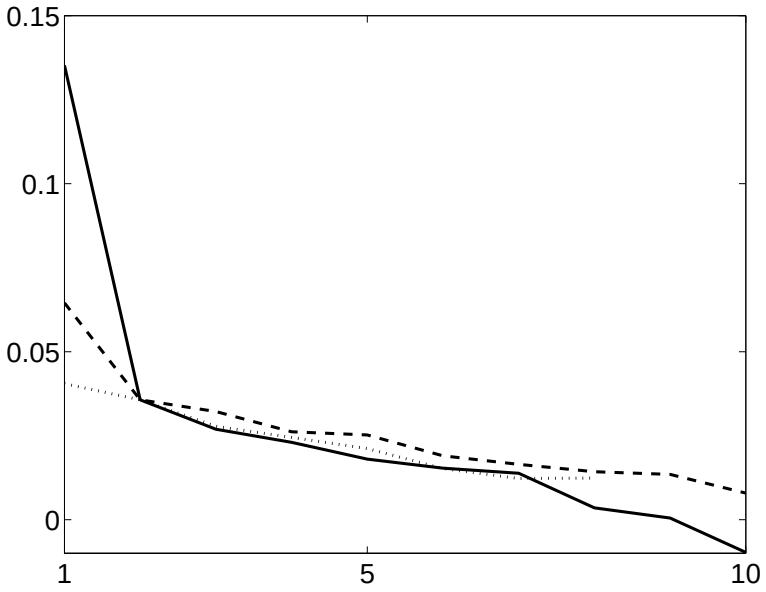


Figure 10: Texture discrimination results. Here the horizontal axis corresponds to the order of the texture pairs in Table 1 and the vertical axis the texture discrimination scores. Dotted line - Psychophysical data from Krose[29]; dash line - results from Malik and Perona's model[32]; solid line - results from the spectral histogram model.

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[32], 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, 47]. Our model may be able to account for the asymmetry phenomenon, and we illustrate this by a simulation involving the commonly used textures of +'s and L's. Fig. 11 shows test patterns used in our simulation. To be consistent with our previous evaluation methodology, we place one texture in the middle and the other one on the two sides. To reflect that the middle one forms the foreground, we compare the peak corresponding to a boundary separating two textures with the peak within the side (background) texture. Note that the layout in Fig. 11 yields two such scores, and the average is taken to indicate the discrimination strength.

For Fig. 11a, the discrimination score produced by our model is 0.005, and for Fig. 11b it is 0.018. In other words, our model predicts that the texture of +'s in the middle of the texture of L's is more difficult to discriminate than when the latter is in the middle of the former. This prediction matches the psychological data[19]. The reason for our model to exhibit discrimination asymmetry is that the variability within the texture of L's is larger than that within the texture of +'s. As a result, the boundary separating the two textures can be relatively stronger or weaker compared to spurious boundaries generated within a background texture. This explanation is similar to that given by Robenstein and Sagi[39], who did a more systematic study on discrimination asymmetry based on a back-pocket model.

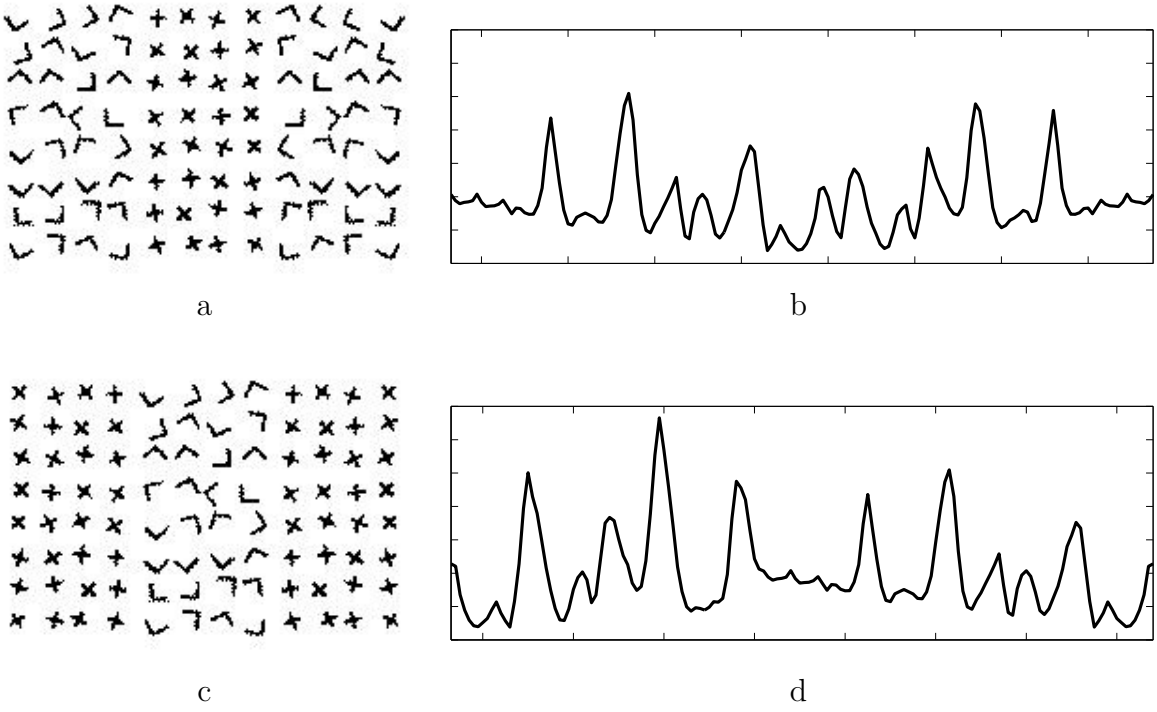


Figure 11: Asymmetry in texture discrimination. a, A texture region of +'s flanked by those of L's. The size of the image is 154×230. b, The average texture gradient of a. The discrimination score produced by the spectral histogram model is 0.005. c, A region of L's flanked by those of +'s. The size of the image is 154×223. d, The average texture gradient of c. The discrimination score is 0.018.

5 Discussion

5.1 Relation to Julesz's Texton Theory

We are directly inspired by the texton theory for our employment and analysis of histograms. According to the Julesz theory[25, 27], preattentive texture segregation occurs between two regions only if they differ in texton density, irrespective of the spatial relationships among textons. A histogram, or marginal distribution, provides a means to represent densities. However, our model builds on spatial/frequency filters and their responses are systematically represented in the spectral histogram. As a result, our model does not need to invoke specialized detectors, which would require a relatively late stage of perceptual computing as well as to limit the scope of the theory (see Sect. 3). Therefore, our model can be applied not only to synthetic textures commonly used in psychophysical experiments but also to natural textures, a point further discussed in Sect. 5.3.

5.2 Relation to Other Studies

Besides the inspiration from the texton theory, our model is related to a number of previous studies. Using a class of independent, identically distributed textures, Chubb, Econopouly, and Landy[9] illustrated that histogram contrast might be used by the visual system to draw distinctions between different image regions.

A histogram analysis is frequently used both for data analysis and image processing [20]. In the literature of texture processing, Unser [42] used the sum and difference histograms of paired pixels with a specified displacement for texture classification, which can be viewed as the spectral histogram of filters whose coefficients are zero everywhere except at two locations with the given displacement. Voorhees and Poggio[43] proposed to employ histograms to compare local distributions of blob detector responses in order to derive texture boundaries. To compare two histograms, they compute the maximum difference between corresponding histogram bins. We have compared their statistic and ours on a systematic database for texture classification, and found that the χ^2-statistic yields better performance[31]. More comparisons with their model are given in Section 5.4. Based on responses from a nonlinear filter, Räth and Morfill[37] used a histogram comparison to derive texture boundaries. Both of these studies were evaluated only on a few images compatible with the employed filters. Ojala et al. [34] compared histograms of a single texture feature and paired texture features for texture classification and showed that the use of distributions of texture features improved classification performance signficantly, in comparison with using texture features directly. Hofmann et al. [23] proposed an objective function for unsupervised texture classification based on a pairwise distance between distributions of Gabor coefficients in local windows. They further performed texture segmentation by minimizing the objective function. In contrast to these studies, which treat a histogram as a texture feature for producing good empirical results, we treat the spectral histogram as an explicit model of texture and verify its validity using extensive texture synthesis.

Heeger and Bergen[21] first studied texture synthesis by matching the histogram pyramids of the observed and the synthesized images, and demonstrated their method on a set of natural textures. Their synthesis technique, however, does not yield an explicit model of

texture and often produces abnormal results[30]. Zhu et al.[49] studied texture synthesis by learning probability models based on distributions of filter responses using statistical sampling. First the probability model is learned based on the observed image(s). Then the texture synthesis is achieved by sampling the learned probability model. While the system is successful in synthesizing natural textures, the learned probability models seem ineffective for synthetic textures; even with texture elements used as filters directly and a specially designed sampling procedure, the synthesized textures are perceptually different from the observed ones[49]. In contrast, our texture model is entirely based on spectral histograms, and it is conceptually consistent with the texton theory. Our improved sampling technique makes it possible to synthesize challenging textures used in psychophysical experiments. Moreover, our texture model has been successfully evaluated on a systematic set of texture discrimination data, while other explicit texture models have not been tested as perceptual models.

5.3 Further Contrast with Back-pocket Models

Sect. 2.2 discusses the relationship between the spectral histogram model and other standard models of texture discrimination. One point not mentioned there concerns natural texture images. Because the visual system normally deals with natural images, a good texture model should, in addition to discriminating synthetic textures, perform well on classifying real textures. Rarely are standard models evaluated on real textures. A main reason for this is that psychophysical experiments almost always use binary, synthetic textures. Such impoverished stimuli are often necessary in controlled human experiments. However, one undesirable consequence is that resulting theories and computational models are often limited to just such stimuli, not applicable to natural images. We think that popular notions such as line, corner, and terminator detectors, in the texture perception literature, even the texton theory itself, are colored by the use of laboratory stimuli.

Randen and Husoy[36] recently performed an extensive evaluation of various texture classification methods that are based on filter responses. Their system setup for comparing purposes includes filtering, nonlinearity, smoothing, and then classification. Thus, the setup may be viewed as a back-pocket model. Their comparisons conclude that no single method performs consistently well on a set of natural images, and this comprehensive study suggests that back-pocket models are inadequate to classify natural textures (see also Chubb and Landy [10]).

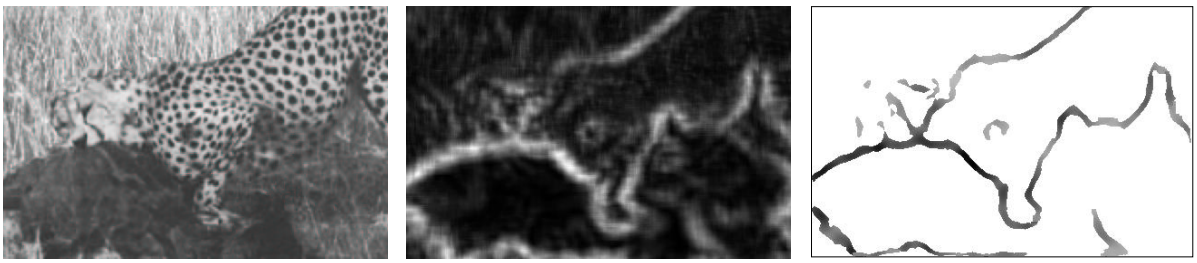
The spectral histogram has been successfully used to classify a large number of real texture images[30, 31]. Our systematic comparison shows that the spectral histogram model substantially outperforms back-pocket models.

5.4 On Texture Segregation

The goal of texture segregation is to produce boundaries separating different texture regions, or, conversely, to segregate an image into regions of homogeneous texture. A potential issue with the spectral histogram model is that, because histogram gathering requires a sizable window, it may lead to grossly inaccurate boundaries. Note that this issue is not unique to our model, and standard models for texture segregation all include a stage of spatial

pooling, which has an effect of blurring boundaries. In essence, texture features require a larger spatial scale to manifest than, say, luminance, color, or motion features.

To illustrate how our model can apply to segmenting natural textures, we process an image that was first used in Voorhees and Poggio[43] and later in Zhu and Yuille[50]. The image, shown in Fig. 12a, contains a cheetah biting a buffalo. Fig. 12b shows the texture gradient of the image produced by our model. The gradient at a pixel location is the sum of the χ^2-distances between the spectral histograms of adjacent windows along a row and a column. In a row or column, the gradient is caculated in the same way as in our texture discrimination experiment. Given the gradient, we detect the texture boundaries by finding local maxima in the gradient image and the resulting texture boundaries are given in Fig. 12c. The cheetah boundaries in Fig. 12c are more extensive and accurate than that generated by Voorhees and Poggio[43]. Ours also yields the boundaries of the buffalo, while theirs does not because their system is specifically designed for blob-like textures. Our boundaries are also more accurate than those produced by Zhu and Yuille.



a b c

Figure 12: Boundary detection for a natural texture image. a, Input image, whose size is 277×422. b, Texture gradient produced by the spectral histogram model. c, Detected texture boundaries.

Fig. 12 is meant to be an illustration. Segmentation of natural textures in the context of spectral histograms is a topic to be dealt with in a separate study (see also Liu[30]), where we suggest a subsequent stage for accurate boundary localization.

5.5 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[8, 12].

To compute a histogram requires neurons with sizable receptive fields, which would presumably occur in the extrastriate cortex[28, 40]. 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 [38, 17]. 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, such as that shown in Fig. 1, 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[15]. Also, evidence from visual perception and neurobiology suggests systematic and concerted temporal coding[44, 33, 6].

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.

6 Conclusion

We have proposed the spectral histogram model as a mathematical definition for a texton, and demonstrated that the model predictions well match a systematic set of texture discrimination data. In addition, the model exhibits both nonlinearity and asymmetry in texture perception. While the present study focuses on textures commonly used in psychophysical experiments, the spectral histogram model performs equally well on natural textures.

In our view, the lack of a quantitative formulation for the texton theory becomes a major obstacle to progress, and our model represents a first attempt towards this end. As such, our model is almost certainly wrong in many aspects. On the other hand, our model provides a solution to several key issues in characterizing elusive textons and in modeling texture discrimination.

Acknowledgments

Authors would like to thank S. C. Zhu for discussions. 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] E. Barth, C. Zetzsche, and I. Rentschler, "Intrinsic two-dimensional features as textons," *J. Opt. Soc. Am. A*, vol. 15, pp. 1723–1732, 1998.

[2] J. Beck, "Perceptual grouping produced by changes in orientation and shape," *Science*, vol. 154, pp. 538–540, 1966.

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

[4] J. R. Bergen, "Theories of visual texture perception," in *Spatial Vision*, D. Regan, ed., CRC Press, Boca Raton, pp. 114–134, 1991.

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

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

[7] T. M. Caelli, "Three processing characteristics of visual texture segmentation," *Spatial Vis.*, vol. 1, pp. 19–30, 1985.

[8] F. W. Campbell and J. G. Robson, "Application of Fourier analysis to the visibility of gratings," *J. Physiol. (Lond.)*, vol. 197, pp. 551–566, 1968.

[9] 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.

[10] C. Chubb and M. S. Landy, "Orthogonal distribution analysis: A new approach to the study of texture perception," in *Computational Models of Visual Processing*, M. S. Landy, and J. A. Movshon, ed., MIT Press, Cambridge, MA, pp. 291–301, 1991.

[11] G. R. Cross and A. K. Jain, "Markov random field texture models," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 5, pp. 25–39, 1983.

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

[13] R. C. Dubes and A. K. Jain, "Random field models in image analysis," *J. Appl. Stat.*, vol. 16, pp. 131–164, 1993.

[14] I. Fogel and D. Sagi, "Gabor filters as texture discriminators," *Biol. Cybern.*, vol. 61, pp. 103–113, 1989.

[15] 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.

[16] 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.

[17] 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.

[18] N. Graham, J. Beck, and A. Sutter, "Nonlinear processes in spatial-frequency channel models of perceived texture segregation: Effects of sign and amount of contrast," *Vision Res.*, vol. 32, pp. 719–743, 1992.

[19] R. Gurnsey and R. A. Browse, "Micropattern properties and presentation conditions influencing visual texture discrimination," *Percept. Psychophys.*, vol. 41, pp. 235–252, 1987.

[20] R. M. Haralick and L. G. Shapiro, "Image segmentation techniques," *Comput. Graphics Image Process.*, vol. 29, pp. 100–132, 1985.

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

[22] E. T. Jaynes, "Information theory and statistical mechanics," *Phys. Rev.*, vol. 106, pp. 620-630, 1957.

[23] T. Hofmann, J. Puzicha, and J. M. Buhmann, "Unsupervised texture segmentation in a deterministic annealing framework," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 20, pp. 803-818, 1998.

[24] B. Julesz, "Visual pattern discrimination," *IRE Trans. Info. Theory*, vol. 8, pp. 84–92, 1962.

[25] B. Julesz, "Textons, the elements of texture perception, and their interactions," *Nature*, vol. 290, pp. 91–97, 1981.

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

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

[28] E. R. Kandel, J. H. Schwartz, and T. M. Jessell, *Principles of Neural Science*, 3rd ed., Elsevier, New York, 1991.

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

[30] X. Liu, "Computational investigation of feature extraction and image organization," Ph.D. Dissertation, Department of Computer and Information Science, The Ohio State University, Columbus, OH, 1999 (available online at http://www.cs.fsu.edu/fsvision/publications/papers/dissertation/index.html).

[31] X. Liu and D. L. Wang, "Texture classification using local spectral histograms," Technical Report OSU-CISRC-7/00-TR17, Department of Computer and Information Science, The Ohio State University, Columbus, OH, 2000 (available online at ftp://ftp.cis.ohio-state.edu/pub/tech-report/2000/TR17.ps.gz).

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

[33] J. W. McClurkin, L. M. Optican, B. J. Richmond, and T. J. Gawne, "Concurrent processing and complexity of temporally encoded neural messages in visual perception," *Science*, vol. 253, pp. 675–677, 1991.

[34] T. Ojala, M. Pietikainen, and D. Harwood, "A comparative study of texture measures with classification based on feature distributions," *Pattern Recognition*, vol. 29, pp. 51-59, 1996.

[35] T. V. Papathomas, C. Chubb, A. Gorea, and E. Kowle, *Early Vision and Beyond*, MIT Press, Cambridge, MA, 1995.

[36] 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.

[37] C. Räth and G. Morfill, "Texture detection and texture discrimination with anisotropic scaling indices," *J. Opt. Soc. Am. A*, vol. 14, pp. 3208–3215, 1997.

[38] K. S. Rockland and J. S. Lund, "Intrinsic laminar lattice connections in primate visual cortex," *J. Comp. Neurol.*, vol. 216, pp. 303–318, 1983.

[39] B. S. Rubenstein and D. Sagi, "Spatial variability as a limiting factor in texture-discrimination tasks: Implications for performance asymmetries," *J. Opt. Soc. Am. A*, vol. 7, pp. 1632–1643, 1990.

[40] 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.

[41] M. R. Turner, "Texture discrimination by Gabor functions," *Biol. Cybern.*, vol. 55, pp. 71–82, 1986.

[42] M. Unser, "Sum and difference histograms for texture classification," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 8, pp. 118-125, 1986.

[43] H. Voorhees and T. Poggio, "Computing texture boundaries from images," *Nature*, vol. 333, pp. 364–367, 1988.

[44] R. J. Watt, "Scanning from coarse to fine spatial scales in the human visual system after the onset of a stimulus," *J. Opt. Soc. Am. A*, vol. 4, 2006–2021, 1987.

[45] R. J. Watt, "Some speculations on the role of texture processing in visual perception," in *Early Vision and Beyond*, T. V. Papathomas, et al., ed., MIT Press, Cambridge, MA, pp. 59–67, 1995.

[46] 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.

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

[48] 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.

[49] 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.

[50] S. C. Zhu and A. Yuille, "Region competition: Unifying snakes, region growing, and Bayes/MDL for multiband image segmentation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 18, pp. 884–900, 1996.