Preattentive texture discrimination with early vision mechanisms

Jitendra Malik and Pietro Perona

Department of Electrical Engineering and Computer Sciences, University of California, Berkeley,
Berkeley, California 94720

We present a model of human preattentive texture perception. This model consists of three stages: (1) convolution of the image with a bank of even-symmetric linear filters followed by half-wave rectification to give a set of responses modeling outputs of V1 simple cells, (2) inhibition, localized in space, within and among the neural-response profiles that results in the suppression of weak responses when there are strong responses at the same or nearby locations, and (3) texture-boundary detection by using wide odd-symmetric mechanisms. Our model can predict the salience of texture boundaries in any arbitrary gray-scale image. A computer implementation of this model has been tested on many of the classic stimuli from psychophysical literature. Quantitative predictions of the degree of discriminability of different texture pairs match well with experimental measurements of discriminability in human observers.

1. INTRODUCTION

Classical theories of texture perception by Julesz[1-3] and Beck[4-6] attribute preattentive texture discrimination to differences in first-order statistics of stimulus features such as orientation, size, and brightness of constituent elements. These theories have typically been constructed for black-and-white dot or line patterns and are not directly applicable to gray-scale images (though Voorhees and Poggio[7] provide a definition of textons for gray-scale images). Experimental results describing phenomena that are not well explained by these theories have been reported.[8-10] An alternative approach[9,11-13,14-16] to texture perception is based on the responses of the linear mechanisms (psychophysically observed spatial-frequency channels and neurophysiologically observed blob-, bar-, and edge-sensitive neurons) that have been used to explain a range of phenomena in early spatial vision. While these efforts have demonstrated that a filtering approach can explain some phenomena that are not consistent with the texton theory, a complete model has not yet been presented. Such a model should satisfy the following criteria:

1. Biological plausibility: The stages of the model should be motivated by, and be consistent with, known physiological mechanisms of early vision.
2. Generality: The model should be general enough that it can be tested on any arbitrary gray-scale image.
3. Quantitative match with psychophysical data: The model should make a quantitative prediction about the salience of the boundary between any two textured regions. Rank ordering of the discriminability of different texture pairs should agree with that measured psychophysically.

We outline our model in Section 2. In Section 3 we motivate the necessity of each stage of the model with physiological, psychophysical, and computational arguments. In Section 4, quantitative predictions from our model are compared with psychophysical data on the discriminability of several texture pairs collected by Gurnsey and Browse[10] and Kröse.[17] We also show in Section 5 that our model performs satisfactorily on the texture pairs invented by Julesz and Kröse[18] and by Williams and Julesz[19] in order to pose difficulties for other quasi-linear-filtering models. We conclude with a critical discussion in Section 6.

2. MODEL FOR TEXTURE PERCEPTION

In the first stage we model the output of V1 simple cells (or subunits of V1 complex cells as described by Spitzer and Hochstein[20]). The image $I(x, y)$ is convolved with a bank of linear filters F_k followed by half-wave rectification. We will indicate the positive part with $R^+(x, y) = \max[R(x, y), 0]$ and the negative part with $R^-(x, y) = \max[-R(x, y), 0]$, which give a set of neural responses $R_i(x, y)$, where the index i identifies the orientation-frequency channel:

$$R_{2k} = (I * F_k)^+(x, y), \qquad R_{2k+1} = (I * F_k)^-(x, y). \qquad (1)$$

Radially symmetric filters model nonoriented simple cells. Directionally tuned filters with even-symmetric cross sections perpendicular to their axes model bar-sensitive simple cells. In Subsection 3.A we give the details of the choice of the filters.

The second stage of our model is nonlinear inhibition, localized in space, within and among the neural-response profiles, which results in the suppression of weak responses when there are stronger responses at the same or nearby locations. Details are in Subsection 3.F. Let $PIR_i(x, y)$ be the postinhibition response in the ith channel.

The third stage of our model is the computation of the texture gradient[21] (Subsection 3.G). We define the texture gradient to be $\max_i \| \nabla (PIR_i * G_{\sigma'})(x, y) \|$, where $G_{\sigma'}$ is a radially symmetric Gaussian function with standard deviation σ' and the index i ranges over all channels.

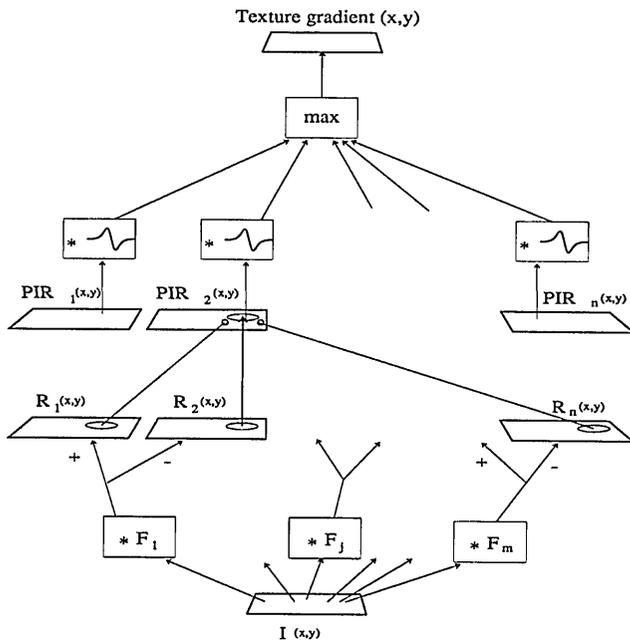A schematic view of the model is presented in Fig. 1.

Fig. 1. Simplified schematics of our model for texture perception. The image (bottom) is filtered using the kernels $F_1 \ldots F_m$ and is half-wave rectified to give the set of simple-cell responses $R_1 \ldots R_n$. The postinhibition responses $\mathrm{PIR}_1 \ldots \mathrm{PIR}_n$ are computed by thresholding the R_i and taking the maximum of the result over small neighborhoods. The thresholds depend on the activity of all channels. The texture gradient is computed by taking the maximum of the responses of wide odd-symmetric filters acting on the postinhibition responses PIR_i.

3. MOTIVATION FOR THE STAGES OF THE MODEL

The general structure of our model follows the findings of Julesz,[1-3] Beck,[4-6] and Treisman[22] that state that in preattentive vision, precise positional relationships between textons are not important; only densities matter. These findings suggest that when two textures T_1 and T_2 are discriminable, they are distinguished by different spatial averages $\int \int_{T_1} R(x, y)$ and $\int \int_{T_2} R(x, y)$ of some locally computed neural response R. A discussion of how earlier models fit into this framework may be found in Ref. 23.

Within this framework, a set of appropriate neural mechanisms that produce responses R_i and a pooling mechanism utilizing these responses to compute the texture gradient have to be chosen. Our guiding principles for these two choices are biological plausibility and parsimony. The final test for the model is, of course, whether it reproduces human performance.

A. Choice of the Filters

Several models have been proposed for the point-spread function of simple cells and subunits of complex cells. These include Gabor functions,[24] differences of offset Gaussians[25] (DOOG), and differences of offset differences of Gaussians.[26] We have chosen to use DOOG (Fig. 2) for our simulations, given their good fit with the physiological measurements and their computational simplicity. We believe that this is not a critical choice. Any of the families of functions mentioned above could have been used instead.

The radially symmetric filter classes $\mathrm{DOG1}(\sigma)$ and $\mathrm{DOG2}(\sigma)$ (Figs. 2a and 2b) model nonoriented simple cells. Directionally tuned filters $\mathrm{DOOG2}$ (σ, r, θ) with even-symmetric cross sections perpendicular to their axes (Fig. 2c) model bar-sensitive simple cells. In our simulations we used six equally spaced orientations θ and a constant aspect ratio $r = 3$.

Implicit in the DOOG model is the assumption that receptive field profiles in the direction that is perpendicular to the axes are either odd-symmetric or even-symmetric and not of an intermediate phase. This model is suggested by psychophysical studies on phase discrimination.[27,28] One has to be aware that electrophysiological mapping of the impulse-response function of single-cortical simple cells does not support this view.[26] At the cell level there seems to be not a sharp dichotomy but rather a continuum between even and odd symmetry. One explanation of this discrepancy could be that the responses of different cells are pooled together in such a way that one effectively gets strictly odd- or even-symmetric mechanisms. We hypothesize (Subsection 3.H) that information from odd-symmetric mechanisms is not used for texture perception and therefore exclude from our model odd-symmetric mechanisms, which respond optimally to appropriately oriented edges.

The σ parameter of the three filter classes that were used corresponds to a nominal spatial frequency in cycles per degree (c/deg) (given the viewing distance and size of image).
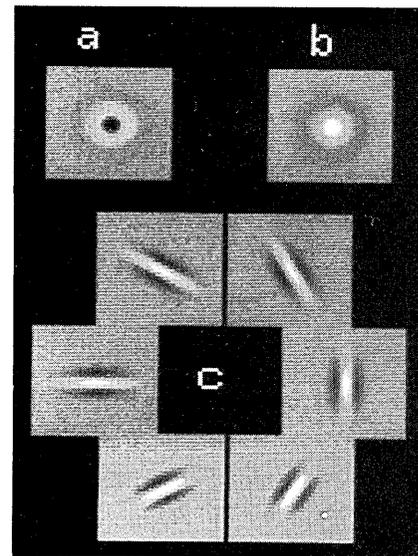

Fig. 2. Point-spread functions of some of the filters used in our simulation. The filters were designed after Young[25] by summing Gaussian functions $G(x_0, y_0, \sigma_x, \sigma_y) \doteq 1/2\pi\sigma_x\sigma_y \exp\{-(x - x_0/\sigma_x)^2 + (y - y_0/\sigma_y)^2]\}$ and have zero-mean value. a, Linear combination of three circular concentric Gaussian functions, $\mathrm{DOG2}(\sigma) \doteq a \cdot G(0, 0, \sigma_i, \sigma_i) + b \cdot G(0, 0, \sigma, \sigma) + c \cdot G(0, 0, \sigma_o, \sigma_o)$ with variance $\sigma_i{:}\sigma{:}\sigma_o$ in a ratio of 0.62:1:1.6 and $a{:}b{:}c$ in a ratio of 1:−2:1. b, Linear combination of two circular concentric Gaussian functions, $\mathrm{DOG1}(\sigma) \doteq a \cdot G(0, 0, \sigma_i, \sigma_i) + b \cdot G(0, 0, \sigma_o, \sigma_o)$, with variance $\sigma_i{:}\sigma_o$ in a ratio of 0.71:1:1.14 and coefficients $a{:}b$ in a ratio of 1:−1. c, Linear combination of three offset identical Gaussian functions $\mathrm{DOOG2}(\sigma, r, \theta) \doteq a \cdot G(0, y_a, \sigma_x, \sigma_y) + b \cdot G(0, y_b, \sigma_x, \sigma_y) + c \cdot G(0, y_c, \sigma_x, \sigma_y)$. Variances are $\sigma_y = \sigma$, $\sigma_x = r \cdot \sigma$, offsets are $y_a = -y_c = \sigma, y_b = 0$, and coefficients are $a{:}b{:}c$ in a ratio of −1:2:−1 for the filter with an axis of symmetry along the x direction ($\theta = 0$). The other $\mathrm{DOOG2}(\,)$ filters are obtained by rotation about the center of the middle Gaussian. The scaling coefficients $a_{\mathrm{DOG1}}{:}a_{\mathrm{DOG2}}{:}a_{\mathrm{DOOG2}}$ were in a ratio of 3:4.15:2, which was designed to equalize the dynamic range of the respective responses.
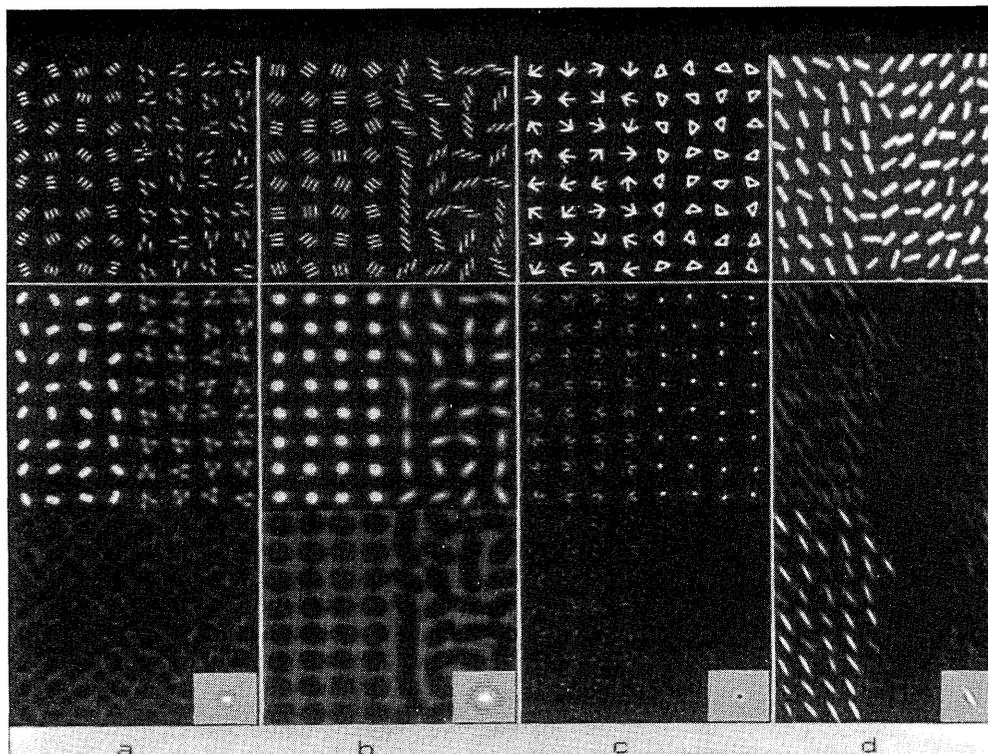
Fig. 3. Some textures (top row) and half-wave-rectified responses of one of the filters to each (bottom row). The point-spread function of each filter is shown at the bottom-right corner of the response image. The filter shapes are as in Fig. 2; the frequency parameters correspond to a 4 deg × 4 deg image. The response images are composed of two square regions, an upper one depicting R^+, the positive part of the response, and a lower one showing R^-. a, Texture from Ref. 10, Fig. 6, pair 2.2 (top) and the response of an 8-c/deg DOG1 filter (bottom); $\sigma \approx 0.5 \times$ (length of texel line segments). b, Texture from Ref. 10, Fig. 6, pair 2.1 (top) and the response of a 5-c/deg DOG1 filter (bottom); $\sigma \approx 2 \times$ (width of texel line segments). c, Arrow–triangle texture (top), for which the arrow texel is obtained from the triangle by shifting one of its legs, and the response to a 5-c/deg DOG2 filter (bottom); $\sigma \approx 0.3 \times$ (length of triangle's hypotenuse). d, Texture from Ref. 30, Fig. 4.2b (top) and the response to a 13-c/deg DOOG2 filter (bottom); $\sigma_y \approx$ (width of bars), $\sigma_x{:}\sigma_y = 3$, and orientation 120 deg.

To sample adequately the spatial-frequency range around the peak of the luminance-contrast-sensitivity function, we used all integer values of the frequency between 3 and 14 c/deg.[29] This gives 96 filters F_k, which result in 192 neural responses R_i. It should be noted that all the filters are zero mean. Consequently, they have zero response to any image region in which the luminance $I(x, y)$ is constant.

Representative examples of these responses for some textures may be found in Fig. 3.

B. Inadequacy of Purely Linear Mechanisms

The mechanisms considered so far—convolutions of the image with V1 cell impulse responses—are linear. To see that a model based purely on linear mechanisms cannot reproduce human performance, we consider two textures T_1, T_2 that have identical mean brightnesses, i.e., identical spatial averages. Convolving them with a linear filter F results in responses $R_{T_1}(x, y)$ and $R_{T_2}(x, y)$ with identical spatial averages. (The values of the power spectra at 0 are identical.) Now, we know that humans can preattentively discriminate some textures with identical spatial averages. An example is the even–odd pair from Ref. 31 or indeed any discriminable texture pair with identical first-order global statistics. A generalization of this observation to nth-order statistics and nth-order polynomial operators may be found in Ref. 23. Some nonlinearity in the system is therefore necessary for texture perception.

The most obvious choice of nonlinearity is half-wave rectification. V1 cortical cells have low-maintained discharge rates and are unable to respond with a decrease in firing rate as required by a negative response. Two different cells are needed (and used) to represent the positive and negative parts of the response belonging respectively to the *on* and *off* pathways.

C. Loss of Essential Information from Full-Wave Rectification

Two nonlinearities that are similar to half-wave rectification have been used in other models of texture discrimination. These are

1. Full-wave rectification, where the response in the kth channel is $R_k(x, y) = |(I * F_k)(x, y)|$, is equivalent to summing the outputs in the two corresponding half-wave rectification channels. This approach has been used by Bergen and Adelson.[14]

2. Energy computation, where $R_k(x, y) = |(I * S_k)(x, y)|^2 + |(I * C_k)(x, y)|^2$ and where S_k, C_k constitute a pair of filters in quadrature phase (e.g., Gabor sine and cosine filters), has been used by Sutter *et al.*[15] and by Fogel and Sagi.[16]

We have two reasons for preferring half-wave rectification. The first is that it is the most natural choice in the context of current biological evidence; we know that linear filtering
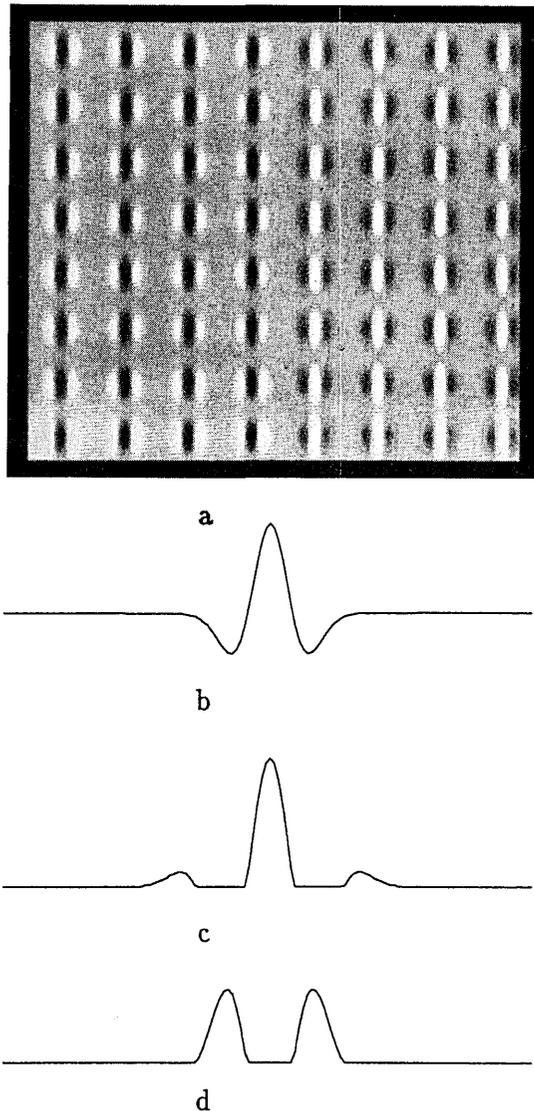
a

b

c

d

Fig. 4.   a, Texture pair that was constructed by adding to a uniform gray field the zero-mean micropatterns M (right) and $-M$ (left). The two textures are easily discriminable, though it may be shown that spatially averaged responses for any linear filter followed by either half- or full-wave rectification are identical for both and thus insufficient for the discrimination. b, Cross section of M along the x axis. c, d, Cross sections of the responses to M and $-M$ in one channel (corresponding to convolution with $F_i = M$ followed by positive half-wave rectification). The areas under c and d are equal. For any zero-mean filter F ($\int \int F = 0$) we have $\int \int (M * F) = 0$; hence $\int \int (M * F)^+ = \int \int (M * F)^- = \int \int (-M * F)^+$.

that is followed by half-wave rectification is a good first-order approximation for modeling responses of simple cells and subunits of complex cells. Cells computing energies have not yet been documented. Second, in both full-wave rectification and energy computation, the sign of the filter response is lost. Consequently, the response in a full-wave rectification channel or in an energy channel is identical for micropatterns M and $-M$ and would remain so after any subsequent processing. This response results in the prediction that a texture pair composed of micropatterns M on one side and $-M$ on the other (bright bars and dark bars, say)

cannot be discriminated. Since we can discriminate easily many such texture pairs (for example, Fig. 4a) it is obvious that these nonlinearities cannot be part of all the channels of the visual pathway that performs texture perception.

D. Need for a Second Nonlinearity

A model of texture discrimination that uses half-wave rectification as the only nonlinearity before the pooling stage (at which the texture gradient is computed) successfully explains human performance on a number of examples. However, we can prove that such a model cannot discriminate texture pairs composed of micropatterns M and $-M$. One such texture pair can be seen in Fig. 4a; human observers have no difficulty in discriminating the bright-bar region from the dark-bar region.

Of course, the responses of a channel (linear filtering followed by half-wave rectification) to micropatterns M and $-M$ are in general going to be different. However when they are pooled, i.e., spatially averaged over a region greater than the area of a micropattern, then they result in identical values. For any zero-d.c. filter $F (\int \int F = 0)$, the average response is $\int \int (M * F) = 0$, which implies that the positive and negative parts of the response on each texture have the same average, i.e., that $\int \int (M * F)^+ = \int \int (M * F)^-$. Now the response to M in the off channel $(M * F)^-$ is the same as that to $-M$ in the on channel $(-M * F)^+$. It follows that $\int \int (M * F)^+ = \int \int (-M * F)^+$ and consequently, pooled responses in any on channel cannot be used to discriminate the two textures. The situation is similar for any off channel. Segmentation is therefore impossible for any channel corresponding to zero-mean filters. For a graphical representation of this argument see Fig. 4.

E. Choices of Second Nonlinearity

The argument in Subsection 3.D provides a motivation for including additional nonlinearities in the channels. However the form and site of the nonlinearity are left largely unspecified.

The first decision to be made is whether to have the nonlinearity precede or follow the simple-cell response stage (linear filtering and half-wave rectification). A well-known early nonlinearity in the visual system—that of retinal adaptation[32]—has been postulated to play a significant role in texture discrimination according to Sutter et al.[15] and Graham et al.[33] This nonlinearity clearly plays an important role in vision because it enables the system to respond to local contrast over ten log units of illumination changes. The crucial question is whether it is sufficient to account for texture discriminability data, with no later nonlinearity required. Relevant evidence comes from data on binary textures with only two distinct gray levels (Section 4), so that local contrasts are equal. For these stimuli, adaptation cannot account for the deviation of rank ordering of psychophysically measured texture discriminability with that predicted in the absence of a late nonlinearity. See the discussion in Section 6. This suggests that a late nonlinearity is essential.

There are at least two physiologically plausible choices for a late nonlinearity:

1. A nonlinear contrast response function g could be picked. Typically g has a sigmoidal shape with neurons that exhibit a threshold effect for low contrast and a saturation

effect for high contrast (e.g., as discussed by Albrecht and Hamilton[34]).

2. Intracortical inhibition could occur within and among the responses in the different channels.

In our model, we have chosen to use intracortical inhibition. We studied a few variations on this theme, which are explored in Subsection 3.F.

F. Nonlinear Inhibition

Inhibitory connections and interactions (some nonlinear) among the neurons in the primary visual cortex have been well documented by physiological and anatomical techniques[35-37] and are presumably the substrate for psychophysically observed inhibition among channels.[38] A number of functional roles, including the generation or sharpening of orientation and length selectivity and contrast gain control, have been attributed to these connections. We propose that another consequence of intracortical inhibition is the suppression or reduction of spurious responses in nonoptimally tuned channels.

What do we mean by spurious responses? First, we discuss some observations about neurons with linear receptive field functions. Such a neuron typically gives a nonzero response to a stimulus to which it is not optimally tuned. For example, consider the texture in Fig. 4a. The bright-bar channel is tuned to the stimulus M (Fig. 4b), and its response (Fig. 4c) has a strong peak at the position of the stimulus. The dark-bar channel is not tuned to the stimulus and gives a response (Fig. 4d) consisting of two smaller peaks. We call this response spurious. The concept is general; the response in an orientationally tuned DOOG2 channel (as in Fig. 2) to a DOG1 stimulus may similarly be regarded as spurious.

In Subsection 3.E we showed an example for which the spatial averaging of the response in the optimal channel gives the same value as the spatial average of a spurious response. The peak value is of course greater for the optimal channel. If intracortical inhibition acted in a way so as to favor responses in optimal channels and reduce the spurious responses, texture discrimination would then be possible by using pooled responses.

Physiological experiments on inhibition in V1 have not yet converged on a definitive model; indeed it is likely that there are several inhibitory circuits with different roles. At present we can only hypothesize a model and argue for (a) its functional adequacy and (b) its biological plausibility. For biological plausibility, we imposed three design constraints on the inhibition model: a neural implementation should (1) require only local connections (in the same or nearby V1 hypercolumns), (2) require only a few neural time steps, and (3) not demand unduly specific interconnection strengths between arbitrary pairs of neurons.

One way to model this inhibition is as follows. Thresholds $T_i(x_0, y_0)$ for neurons belonging to channel i with retinotopic coordinates x_0, y_0 are computed:

$$T_i(x_0, y_0) = \max_j \max_{x,y \in I_{ji}(x_0,y_0)} \alpha_{ji} R_j(x, y). \quad (2)$$

Here I_{ji} is the neighborhood of (x_0, y_0) in which neurons in channel j are able to inhibit neurons in channel i, and α_{ji} is a measure of the effectiveness of this inhibition. The postinhibition response $PIR_i(x_0, y_0)$ is given by

$$PIR_i(x_0, y_0) = \max_{x,y \in S_i(x_0,y_0)} \frac{1}{1 - \alpha_{ii}} [R_i(x, y) - T_i(x, y)]^+. \quad (3)$$

This results in a suppression of responses below the threshold. $S_i(x_0, y_0)$ is a sampling neighborhood from which the strong responses in channel i are selected for subsequent processing.

One way to think about this mechanism is as a so-called leaders-take-most feedforward network; it is a variant of the winner-take-all type mechanisms quite popular in the neural-network literature.

We were guided in our choice of I_{ji} and α_{ji} by the design criterion of eliminating spurious responses. As an example, consider channels i and j that correspond to the positive and negative responses of the filter $DOOG2(\sigma)$. Figures 4c and 4d show the responses in the two channels to a bright bar. The peaks in the negative (spurious) response are approximately 0.65 times the positive central peak and are displaced from it by 1.25σ. This prompts a choice of $\alpha_{ji} = 0.65$ and I_{ji} to be a disk of radius 1.25σ in order to ensure a suppression of the negative response. This procedure can be repeated for all the 192×192 pairs of channels. However this violates our third criterion for biological plausibility because specific interconnection strengths are required between arbitrary pairs of neurons. We can however exploit the known clustering; nonorientationally tuned neurons tend to occur in the V1 blobs, and neurons sharing similar orientation preferences occur together. This clustering leads us to form eight groups of channels in our framework (two radially symmetric + six oriented). I_{ji} and α_{ji} are identical for all channels i in one of these groups; these values have been computed from the spurious responses in the channel i with the same σ parameter as channel j. The actual values used in our simulations are shown in Tables 1 and 2. S_i was chosen to be a disk of constant radius. We will refer to this model of inhibition as model A.

Model A seems rather elaborate, and it is natural to seek simpler models that might be adequate. In model B, we set $\alpha_{ji} = 0.5$. This corresponds to having a nonspecific local inhibitory pool of neurons and would be more consistent with the physiological evidence of Bonds.[39] In model C, α_{ji}

Table 1. Inhibitory Coefficients $\alpha_{ji}{}^A$ for Model A[a]

i	j		
	$DOG1(\sigma_j)$	$DOG2(\sigma_j)$	$DOOG2(\sigma_j, r_j, \theta_j)$
$DOG1(\sigma_i)$	0.2	0.45	0.15
$DOG2(\sigma_i)$	0.45	0.25	0.20
$DOOG2(\sigma_i, r_i, \theta_i)$	0.15	0.20	$0.65 \, \delta(\theta_i, \theta_j)$

[a] $\delta(\theta_i, \theta_j) = 1$ if $\theta_i = \theta_j$; 0 otherwise. The inhibitory coefficients for models B, C, and D were constant: $\alpha_{ji}{}^B \equiv 0.5$, $\alpha_{ji}{}^C \equiv 0$, and $\alpha_{ji}{}^D \equiv 0$.

Table 2. Radii of the Inhibition Neighborhoods I_{ji} for Models A and B[a]

i	j		
	$DOG1(\sigma_j)$	$DOG2(\sigma_j)$	$DOOG2(\sigma_j, r_j, \theta_j)$
$DOG1(\sigma_i)$	$2\sigma_j$	$1.5\sigma_j$	$1.25\sigma_j$
$DOG2(\sigma_i)$	$2\sigma_j$	$1.5\sigma_j$	$1.25\sigma_j$
$DOOG2(\sigma_i, r_i, \theta_i)$	$2\sigma_j$	$1.5\sigma_j$	$1.25\sigma_j$

[a] In Models C and D the neighborhoods I_{ji} are irrelevant.

= 0, which eliminates the threshold but keeps the stage of computing local, strong responses in a neighborhood S_i. One could relate this model to outputs of complex cells that do not show a significant modulation in their response when a stimulus is moved across the receptive field. Finally, model D served as a control with $PIR_i = R_i$.

G. Computation of the Texture Gradient

Nothdurft[21] has shown two characteristics of texture perception that any model must explain. Texture discriminability depends on (a) the density of micropatterns in the image, with higher densities leading to easier discrimination, and (b) local differences rather than global differences. This naturally suggests the idea of computing the gradient of the smoothed postinhibition responses in each channel.

The texture gradient that we use is defined as $\max_i \nabla$ $(PIR_i * G_{\sigma'})(x, y)$. Biologically, the computation of the gradient of the smoothed postinhibition response in each channel can be done by using odd-symmetric oriented mechanisms similar to the edge-sensitive cells in V1. Of course, the mechanisms responsible for computing the texture gradient have large receptive fields (σ' is a measure of the size) and presumably occur in some extrastriate area. The maximum operation seems a natural way of combining the outputs of the different channels. Texture boundaries may be defined as corresponding to local peaks of the texture gradient magnitude (Fig. 5).

H. Nonuse of Odd-Symmetric Mechanisms

Our model used only channels corresponding to even-symmetric filters. This choice was based on an interpretation of some experimental results of Rentschler et al.,[40] who found that textures composed of mirror-image, compound Gabor signals were indistinguishable even when the individual micropatterns were easily discriminated. There was no difficulty in discriminating textures composed of nonmirror-image, compound Gabor signals. A simplified version of the phenomenon can be seen by comparing Fig. 4a (easily segmentable) and Fig. 6 (not preattentively segmentable). We will show that this phenomenon implies that odd-symmetric and even-symmetric filters are not treated identically in texture discrimination. Specifically, the signs of responses of odd-symmetric filters are ignored, while the signs of the responses of even-symmetric filters are used (for example, to distinguish dark-bar and bright-bar textures as in Fig. 4).

First we supply some definitions: micropatterns M_1 and M_2 are said to be y mirror symmetric (y-ms) if $M_1(x) = M_2(-x)$ and xy mirror symmetric (xy-ms) if $M_1(x) =$
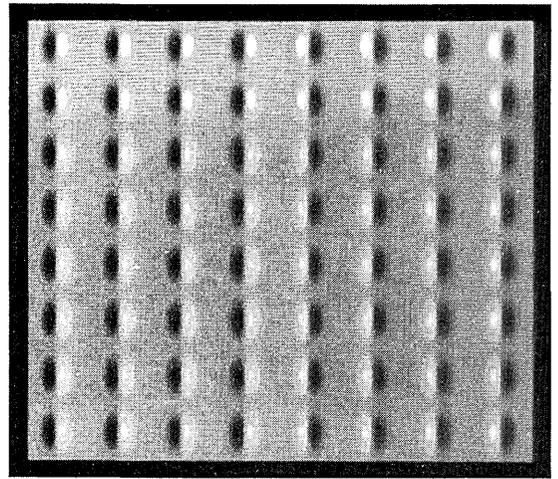
Fig. 6.   Texture pair composed of y mirror-symmetric micropatterns. Segmentation is not preattentive. Compare with Fig. 4.

$-M_2(-x)$. Examples of y-ms pairs are found in Ref. 40 (Figs. 2a and 2c) and in the two micropatterns in Fig. 6; Fig. 4 contains an xy-ms pair. Consider any two y-ms patterns M_1, M_2. Now, the following operations (or any composition thereof) preserve y-ms: (a) half-wave rectification, (b) convolution with any even-symmetric filter, and (c) nonlinear scaling $I \rightarrow g(I)$. Consequently, responses $R_i(M_1), R_i(M_2)$ in any channel i corresponding to an even-symmetric filter are also y-ms. In fact, so also are postinhibition responses $PIR_i(M_i)$, $PIR_i(M_2)$ if only inhibition from channels j corresponding to even-symmetric filters is considered [for any such j, $R_j(M_1)$, $R_j(M_2)$ are y-ms, resulting in $T_i(M_1)$, $T_i(M_2)$, the respective thresholds being a y-ms pair]. Now any two patterns that are a y-ms pair have identical spatial averages, and from the preceding argument so must postinhibition responses in even-symmetric channels. In other words, to segment a texture composed of M_1 from one composed of M_2 by using spatially averaged responses, we must rely on the channels corresponding to odd-symmetric filters. Interestingly, for an xy-ms pair, the situation is reversed; only even-symmetric filters are useful. To establish this, note that convolving an xy-ms pair with an odd filter makes it a y-ms pair.

To find the texture boundary in Fig. 6, the visual system must rely on the differential activation of channels corresponding to odd-symmetric filters; the detection of texture boundary in Fig. 4a relies on even-symmetric channels. The latter is easily discriminable; the one in Fig. 6 is not. One could conclude from this result that odd-symmetric mechanisms are not utilized in texture perception but that even-symmetric are. This could be because (a) odd-symmetric mechanisms are not part of the texture processing pathway or (b) inhibitory interactions between odd-symmetric cells are such that their activity is greatly reduced when they are stimulated by repetitive texture patterns.

An alternative hypothesis is that the outputs of odd-symmetric cells of opposite polarities are pooled together in the texture-processing pathway, and therefore the information necessary for segmenting y-ms textures is lost.

We are not in a position to discriminate precisely among these hypotheses. Since we have not found any textures for
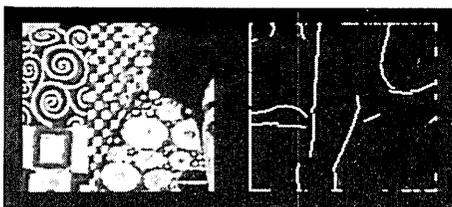

Fig. 5.   Detail of the portrait of Adele Bloch–Bauer by Gustav Klimt (left) and the texture boundaries that were found (right). The essential boundaries of the five perceived groups have been detected.

which odd-symmetric mechanisms are necessary, we have chosen to exclude odd-symmetric mechanisms from our model.

4. EXPERIMENTAL RESULTS

We have compared the degree of texture discriminability that was predicted by our algorithm with psychophysical data from Kröse[17] and Gurnsey and Browse.[10] Figure 7 shows seven bipartite textures with elements constructed after Kröse[17] (Section 3.2, pp. 34–39), two after Williams and Julesz,[19] and one composed of R's and mirror-image R's (called R-mirror-R). For two of these textures, the texture gradient ($\sigma' = 12$ pixels, $S_i = $ constant) obtained by our algorithm, using model A for inhibition, is plotted as a function of column number (Fig. 8). The texture boundary (column 64) is associated with the central peak in the gradient. The value of the gradient associated with this peak is taken to be a measure of the discriminability predicted by our algorithm. In Table 3, these data are presented in a more easily readable form and compared with data from Kröse (Table 3.1, p. 39; stimulus onset asychrony, 320) and Gurnsey and Browse (pairs 1.1, 1.2, 1.3, 3.1) for mean overall discriminability. Note that the rank order of discriminability predicted by our model matches the rankings found experimentally.

The Williams–Julesz textures were constructed to demonstrate essential nonlinearities in texture preception. Their
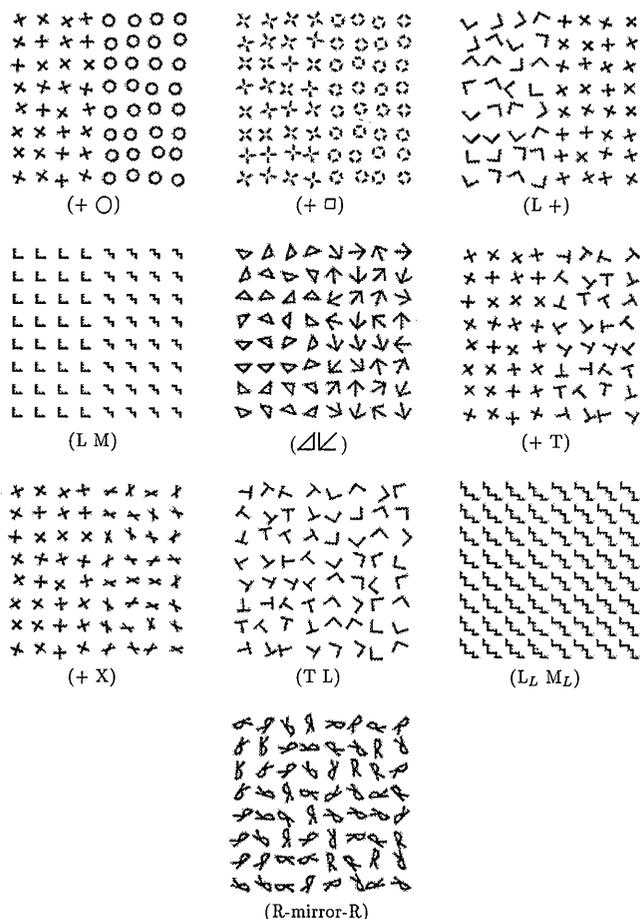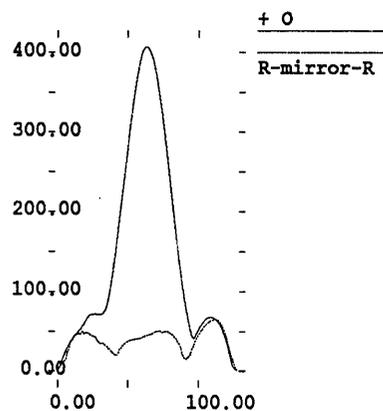

Fig. 8. Texture gradient as a function of column number. For the 128 × 128 textures in Fig. 7 the texture gradient is averaged along the vertical direction on the central middle portion of each column and plotted with respect to the horizontal coordinate. Such plots are shown for the most (L +) and least (R-mirror-R) discriminable textures. The value of the texture gradient at its central peak is taken to be the prediction of our model and is reported in Table 3, column 3.

Table 3. Comparison of Predictions from Texture Segmentation Algorithm with Two Sets of Psychophysical Data[a]

| Texture Pair | Discriminability | | |
	Data Refs. 41 and 42	Data Ref. 43	Predicted Data
+ O	100(saturated)	n.a.	407
+ □	88.1	n.a.	225
L +	68.6	0.736	203
L M	n.a.	n.a.	165
$\Delta^{\cancel{}}$	52.3	0.4–0.55	159
+ T	37.6	0.496	120
+ X	30.3	n.a.	104
T L	30.6	0.421	90*
$L_L\, M_L$	n.a.	n.a.	85
R-mirror-R	n.a.	n.a.	50*

[a] The symbol * indicates that a side peak of the texture gradient was higher than the reported central peak. Because of differences in the scales used, the three columns should be compared only by the rank ordering of discriminability. The rank order of discriminability for the predicted data matches both other data rankings. The L M and $L_L\, L_L$ textures have been invented by Williams and Julesz as a counterexample to purely linear theories.[19] Our algorithm correctly ranks the L M pair within the most discriminable textures and the $L_L\, M_L$ pair within the least discriminable ones. The discriminability of the + O texture given by Kröse saturates his psychophysical scale (top value, zero standard deviation), so it cannot be compared quantitatively with the other discriminability figures (standard deviation ranging between 6.7 and 11.7); n.a., not available. Also compare Fig. 8.

(+ O)

(+ □)

(L +)

(L M)

(△⫿)

(+ T)

(+ X)

(T L)

($L_L\, M_L$)

(R-mirror-R)

Fig. 7. Nine textures that were used in our experiments.

reasoning is as follows. The *LM* texture is easily discriminable; not so the $L_L M_L$ texture (call it C), which is obtained by adding to the *LM* texture (call it A) a uniform texture of little L's (call it B) placed at the endpoints of the L and M micropatterns. If the discriminability between the left and right regions were a linear function of the image, then the discriminability of C = A + B would be the sum of the discriminabilities of A and B. As B is a single-texture region its discriminability is of course zero, so the discriminability of C should be equal to that of A. Clearly it is not.

The match with the experimental data that we have obtained is surprisingly good; we are not aware of any other

Table 4. Comparison of the Predictions from Models A–D with Segmentability Measurements for Two Sets of Experimental Data[a]

			Discriminability			
				Predicted Data		
Texture Pair	Data Refs. 41 and 42	Data Ref. 43	Model A Specific Inhibition	Model B Constant Inhibition	Model C No Inhibition	Model D No Inhibition, S_i = 1 pixel
+ O	100(sat)	n.a.	407	450	235	35
+ □	88.1	n.a.	225	306	140	8*
L +	68.6	0.736	203	327	202	20
Δ⌐	52.3	0.4–0.55	159	172	112	12
+ T	37.6	0.496	120	189	126	12*
+X	30.3	n.a.	104	170	158	10*
T L	30.6	0.421	90*	170	87*	8*

[a] The symbol * indicates that a side peak of the texture gradient was higher than the central peak reported.

Table 5. Comparison of the Discriminability Ranking Given by Models A–D with That for Experimental Data

		Predicted Data			
Ranking	Data Refs. 41 and 42	Model A	Model B	Model C	Model D
1	+ O	+ O	+ O	+ O	+ O
2	+ □	+ □	+ L	L +	L +
3	L +	L +	+ □	+ X	+T, Δ⌐
4	Δ⌐	Δ⌐	+ T	+ □	
5	+ T	+ T	Δ⌐, T L, + X	+ T	+ X
6	+ X, T L	+ X		Δ⌐	T L, + □
7		T L		T L	

model that fits these data. Of course, the usual notes of caution for any model with parameters that are not directly measured from physiology or psychophysics apply. The particular equations and parameters that we have proposed are surely wrong in detail. To have any relevance to biological texture perception, the model should degrade gracefully, i.e., roughly similar ideas should work as well, and choices of parameters should not be too critical.

One can gain additional insight into these issues by studying simplified variants of the inhibition model A. We did that by replacing it with the models B, C, and D described in Subsection 3.F. The results of these simulations are shown in Tables 4 and 5. Additional simulation results with a slightly different choice of parameters of model A may be found in Ref. 44.

Some qualitative features of the results are noted below. Appropriate caution should be exercized as we do not have any error bars for the simulation data for the four models.

1. Model A matches the experimental data best.
2. Model B does almost as well. The order of discriminability of the textures (+ □) and (L +) is reversed. Most significant is the decrease in the discriminability of the texture (Δ⌐); it is now in the same group as the three significantly less discriminable textures (+ T), (T L), and (+ X).
3. For model C, we note that, in addition to the errors mentioned for model B, we have the unwanted increase in the discriminability of the texture (+ X).
4. Model D serves as a control, with $PIR_i = R_i$. The errors here can not be explained by the adaptation nonlinearity proposed by Graham et al.[33] For these binary images, contrast is the same for all micropatterns.

5. LAPLACIAN PYRAMID TEXTURES

Julesz and Kröse[18] studied a texture, which was composed of L's and +'s, that Bergen and Adelson[14] say is segmentable by using the output of a simple center-surround Laplacian-of-Gaussian filter. Julesz and Kröse decomposed this image into a series of bandpass-filtered images by using the Laplacian pyramid technique. Levels 2, 3, and 4 of the pyramid appeared to be the ones showing a difference between the texture of L's and the texture of T's and hence responsible for texture discrimination. Julesz and Kröse constructed a new image for which these levels had been replaced by uniform gray. The textures however remained highly discriminable, which casts doubt on the validity of Bergen and Adelson's explanation.

We re-created this phenomenon (Fig. 9). The three textures appear equally discriminable, and a correct model of texture perception should be able to predict this fact. We tested our model on these textures, and the results may be found in Fig. 10. It is seen that our model finds the textures approximately equally discriminable, as indeed it should. On examining various bandpass-filtered versions of the textures, we observed that at most frequency bands there is some difference between the L's texture and the +'s texture; this difference can be amplified and used by the subsequent nonlinear stages of the model. In our simulation we found the most significant differences in the postinhibition responses PIR_i correspond to the positive part of DOG1 filters between 3 and 9 c/deg for the original +−L texture, corresponding indeed to the levels 2, 3, and 4 of the Laplacian pyramid. For the +−L texture deprived of the levels 2, 3, and 4 of the Laplacian pyramid, the DOG2 filters with fre-

quencies above 10 c/deg were the most important for segmentation. The shapes of the filters are specified in Fig. 2.

Julesz[45] has suggested that the technique of removing frequency bands that was used by Julesz and Kröse can be used to generate counterexamples for theories that do not have any nonlinearity before the linear filtering stage. We have tried this on the +−L and Δ^j textures by deleting the frequencies that appeared to be used by our algorithm for segmentation. A segmentable texture obtained this way would falsify our theory or, at least, show that we need to add more filters to our implementation. As observed above, for the +−L texture every frequency band appears to contribute to segmentation and so produces the trivial result that only a blank image is not segmentable. In the Δ^j texture, by filtering away the relevant frequency bands we considerably
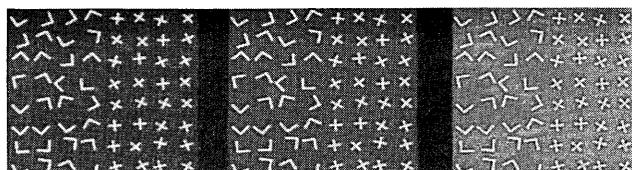

Fig. 11. Texture (Δ^j) (center) and two textures obtained by its bandpass filtering (left and right). Our simulations suggested that segmentation of Texture (Δ^j) does not rely on outputs of filters of 3–4, 10–11, and 15 c/deg. The two bandpassed images were obtained by filtering the original texture with radially symmetric window (brick wall) filters passing the corresponding bands (left) and a comparable number of complementary bands (right). Our model predicts that the left image should be much less segmentable than the one to the right. The bands used were (3–9 30–42 48–57) for the left image and (9–30 42–48) for the right image (frequency units here are in pixels and refer to 128 × 128 pixel-square images, which is not to be confused with the c/deg units that are used in the rest of this paper).

reduce discriminability (see Fig. 11), as expected from our model.

6. DISCUSSION

The results in Section 4 illustrate the explanatory power of our model and suggest that many of the essential aspects of texture perception have been captured in our theory. We list here the principal contributions:

1. Theoretical arguments pointing out the need for essential nonlinearities in texture perception and critiquing full-wave rectification, energy computation, and adaptation as choices of these nonlinearities.

2. Arguments pointing out that the polarities of responses of odd-symmetric mechanisms are not utilized in texture discrimination.

3. A demonstration that a model using half-wave rectification and nonlinear inhibition can explain psychophysical data on degree of texture discriminability. (This model works equally well on the tricky examples of Julesz and Kröse as well as Williams and Julesz.)

Weaknesses of our model, which suggest further research, include the following list:

1. Our model of nonlinear inhibition is somewhat *ad hoc*, as it was constructed in the absence of detailed quantitative experimental data. We hope that further research in physiology and psychophysics will provide these data. The precise form of this stage (subthreshold suppression) in our model should be viewed as a stopgap approximation.

2. Julesz,[1] Treisman,[22] and Gurnsey and Browse[10] have all mentioned the asymmetries in various tasks in preattentive vision. It is most clearly seen in the data from Gurnsey and Browse, e.g., when a field of L's is embedded in a field of +'s, its discriminability (0.93) is much greater than for a field of +'s embedded in L's (0.53). We suspect that this phenomenon is related to noise in the texture gradient. Similar ideas in the context of a different model of texture discrimination have been expressed by Rubenstein and Sagi.[46]

Fig. 9. Textures from left to right: the +L texture, the same after removal of Laplacian pyramid level 3 (+L −3), the same after removal of Laplacian pyramid levels 2, 3, and 4 (+L −234). The Laplacian pyramid was generated by taking differences of contiguous levels of a Gaussian pyramid. Level 0 of the Gaussian pyramid was the image itself; level i was the image convolved with a rotationally symmetric Gaussian of unitary norm, and σ is equal to 2^i pixels. The original image is 128 × 128 pixels in size. In our experiment the (+ L) image and the (+ L −3) image were scaled by $2/3$ and $4/5$, respectively, to reach roughly the same perceptual segmentability as the (+L − 234) image.
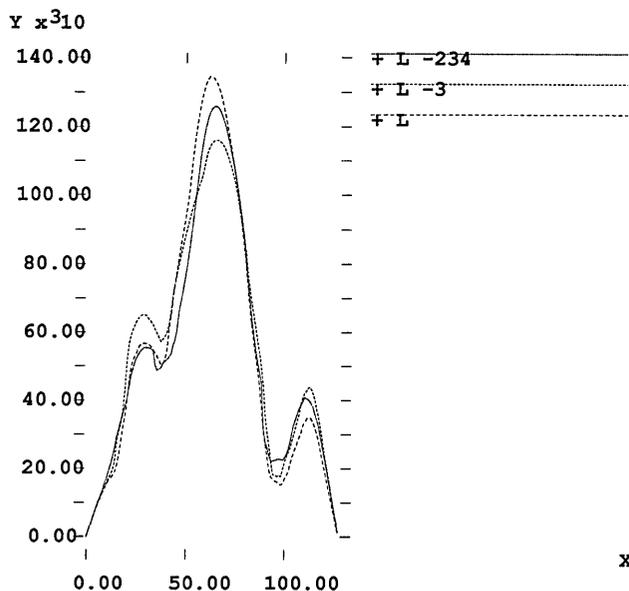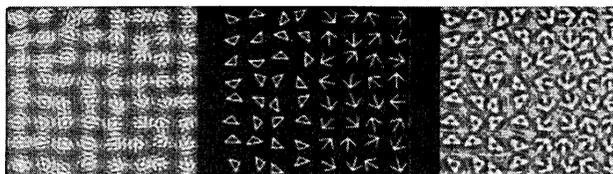

Fig. 10. Texture gradient for the three textures in Fig. 9. The average gradient over the central middle portion of each column of the picture is plotted. The values of the maxima are 134 for the (+ L) image, 114 for the (+ L −3) image, and 126 for the (+ L −234) image. These values have to be scaled by $3/2$ to be compared with the values in Fig. 8.

3.   In our simulations, the choice of σ' was made by hand (8, 12, or 16 pixels) in the computation of the texture gradient.

ACKNOWLEDGMENTS

REFERENCES AND NOTES

1. B. Julesz, "Textons, the elements of texture perception and their interactions," Nature (London) 290, 91–97 (1981).
2. J. Bergen and B. Julesz, "Rapid discrimination of visual patterns," IEEE Trans. Syst. Man Cybern. 13, 857–863 (1983).
3. B. Julesz, "Texton gradients: the texton theory revisited," Biol. Cybern. 54, 245–251 (1986).
4. J. Beck, "Similarity grouping and peripheral discriminability under uncertainty," Am. J. Psychol. 85, 1–19 (1972).
5. J. Beck, "Textural segmentation," in Organization and Representation in Perception, J. Beck, ed. (Erlbaum, Hillsdale, N.J., 1982).
6. J. Beck, K. Prazdny, and A. Rosenfeld, Human and Machine Vision (Academic, New York, 1983), pp. 1–38.
7. H. Voorhees and T. Poggio, "Computing texture boundaries from images," Nature (London) 333, 364–367 (1988).
8. J. Enns, "Seeing textons in context," Percept. Psychophys. 39, 143–147 (1986).
9. J. Beck, A. Sutter, and R. Ivry, "Spatial frequency channels and perceptual grouping in texture segmentation," Comput. Vision Graphics Image Process. 37, 299–325 (1987).
10. R. Gurnsey and R. Browse, "Micropattern properties and presentation conditions influencing visual texture discrimination," Percept. Psychophys. 41, 239–252 (1987).
11. T. Caelli, "Three processing characteristics of visual texture segmentation," Spatial Vision 1, 19–30 (1985).
12. J. Coggins and A. K. Jain, "A spatial filtering approach to texture analysis," Pattern Recogn. Lett. 3, 195–203 (1985).
13. M. Turner, "Texture discrimination by gabor functions," Biol. Cybern. 55, 71–82 (1986).
14. J. Bergen and E. Adelson, "Early vision and texture perception," Nature (London) 333, 363–364 (1988).
15. A. Sutter, J. Beck, and N. Graham, "Contrast and spatial variables in texture segregation: testing a simple spatial-frequency channels model," Percept. Psychophys. 46, 312–332 (1989).
16. I. Fogel and D. Sagi, "Gabor filters as texture discriminators," Biol. Cybern. 61, 103–113 (1989).
17. B. J. Kröse, A Description of Visual Structure, Ph.D. dissertation (Delft University of Technology, Delft, The Netherlands, 1986).
18. B. Julesz and B. Kröse, "Features and spatial filters," Nature (London) 333, 302–303 (1988).
19. B. Julesz, AT&T Bell Laboratories, Murray Hill, New Jersey 07974 (personal communication).
20. H. Spitzer and S. Hochstein, "Simple- and complex-cell response dependences on stimulation parameters, and A complex cell receptive-field model," J. Neurophysiol. 53, 1244–1286 (1985).
21. H. C. Nothdurft, "Sensitivity for structure gradient for texture discrimination tasks," Vision Res. 25, 1957–1968 (1985).
22. A. Treisman, "Preattentive processing in vision," Comput. Vision Graphics Image Process. 31, 156–177 (1985).
23. P. R. Kube, On Image Texture, Ph.D. dissertation (University of California, Berkeley, Berkeley, Calif., 1988).
24. J. D. Daugman, "Two dimensional spectral analysis of cortical receptive field profiles," Vision Res. 20, 847–856 (1980).
25. R. Young, "The Gaussian derivative theory of spatial vision: analysis of cortical cell receptive field line-weighting profiles," Tech. Rep. GMR-4920 (General Motors Research, Warren, Mich., 1985).
26. A. Parker and M. J. Hawken, "Two-dimensional spatial structure of receptive fields in monkey striate cortex," J. Opt. Soc. Am. A 5, 598–605 (1988).
27. D. Field and J. Nachmias, "Phase reversal discrimination," J. Vis. Res. 24, 333–340 (1984).
28. D. Burr, C. Morrone, and D. Spinelli, "Evidence of edge and bar detectors in human vision," Vision Res. 29, 419–431 (1989).
29. We have used a linear sampling of the frequency space instead of the more common logarithmic sampling. The way we combine the output of the different channels makes this choice immaterial, provided that the sampling is dense enough.
30. H. Voorhees, "Finding texture boundaries in images," Tech. Rep. 968 (Massachusetts Institute of Technology, Artificial Intelligence Laboratory, Cambridge, Mass., 1987).
31. B. Julesz, E. N. Gilbert, and J. D. Victor, "Visual discrimination of textures with identical third order statistics," Biol. Cybern. 31, 137–140 (1978).
32. R. Shapley and C. Enroth-Cugell, "Visual adaptation and retinal gain controls," Prog. Retinal Res. 4, 263–347 (1984).
33. N. Graham, J. Beck, and A. Sutter, "Two nonlinearities in texture segregation," Invest. Ophtalmol. Vis. Sci. 30, 161 (1989).
34. D. Albrecht and D. Hamilton, "Striate cortex of monkey and cat: contrast response function," J. Neurophysiol. 48, 217–237 (1982).
35. K. Toyama, M. Kimura, and K. Tanaka, "Organization of cat visual cortex as investigated by cross-correlation techniques," J. Neurophysiol. 46, 202–214 (1981).
36. K. De Valois and R. Tootell, "Spatial-frequency-specific inhibition in car striate cortex cells," J. Physiol. 336, 359–376 (1983).
37. A. M. Sillito and P. C. Murphy, Neurotransmitters and Cortical Function: From Molecules to Mind (Plenum, New York, 1988), Chap. 11.
38. D. Tolhurst, "Adaptation to square wave gratings: inhibition between spatial frequency channels in the human visual system," J. Physiol. 226, 231–248 (1972).
39. A. B. Bonds, "Role of inhibition in the specification of orientation selectivity of cells in the car striate cortex," Visual Neurosci. 2, 41–55 (1989).
40. I. Rentschler, M. Hubner, and T. Caelli, "On the discrimination of compound Gabor signals and textures," Vision Res. 28, 279–291 (1988).
41. Data are from Ref. 17. The tabulated data correspond to δ_{tb} (Table 3.1, p. 39; stimulus onset asynchrony, 320).
42. B. Kröse, "Local structure analyzers as determinants of preattentive pattern discrimination," Biol. Cybern. 55, 289–298 (1987).
43. Data are from Ref. 10. The tabulated data correspond to mean overall discriminability (pairs 1.1, 1.2, 1.3, 3.1) averaged over foreground/background and different stimulus durations.
44. J. Malik and P. Perona, "A computational model of texture perception," Tech. Rep. UCB/CSD 89/491 (Computer Science Division, University of California, Berkeley, Berkeley, Calif., 1989).
45. B. Julesz, AT&T Bell Laboratories, Murray Hill, New Jersey 07974 (personal communication).
46. B. Rubenstein and D. Sagi, "Texture variability across the orientation spectrum can yield asymmetry in texture discrimination," Perception 18, 517 (1989).
47. J. Malik and P. Perona, "A computational model of human texture perception," Invest. Ophthalmol. Vis. Sci. 30, 161 (1989).