A computational model of texture segmentation

Jitendra Malik and Pietro Perona

Department of Electrical Engineering and Computer Sciences
University of California, Berkeley, CA 94720
e-mail: malik@ernie.berkeley.edu

Abstract

We have developed an algorithm for finding texture boundaries in images based on a computational model of human texture perception. The model consists of three stages: (1) The image is convolved with a bank of even-symmetric linear filters followed by half wave rectification to give a set of responses (models outputs of V1 simple cells). (2) Inhibition, localized in space, within and among the neural response profiles which results in the suppression of weak responses when there are strong responses at the same or nearby location (models intracortical inhibition in V1), and (3) texture boundary detection using peaks in the gradients of the inhibited response profiles. Unlike previous attempts along these lines, our model is precisely specified, equally applicable to grey scale and binary textures, and is motivated by detailed comparison with psychophysics and physiology. This model has been tested on a large number of the 'classic' stimuli from psychophysical literature. Our model makes predictions about the degree of discriminability of different texture pairs which match very well with experimental measurements of discriminability in human observers due to Kröse and Gurnsey & Browse.From a machine vision point of view our scheme is a high quality texture edge detector which works equally well on images of artificial and natural scenes. The algorithm makes use of simple, local and parallel operations which makes it potentially realtime.

1 Introduction

Detecting texture boundaries in images is a very important problem in machine vision. This is because most perceptually useful boundaries are between differently textured surfaces. The usual technique in the machine vision community is to detect step edges in image brightness at various scales, and hope that at one of the coarse scales the texture boundary will correspond also to a brightness edge. Clearly, this is not a robust method. Image brightness step edges are valid abstractions only at the finest level of detail, and except for shadow boundaries, step edges at coarser scales are physically meaningful only when they happen to coincide with texture boundaries.

There are very few algorithms in the literature for finding texture boundaries in real images in the absence of prior models of the specific textures present. Two recent efforts are due to Voorhees and Poggio[31,32], and Kashyap and Eom [19]. Surveys of earlier work on texture, mainly focussing on classification, may be found in Haralick [16],and Van Gool, Dewaele and Oosterlinck [30].

Voorhees and Poggio[31,32] compute 'blobs' by convolving the image with a Laplacian of Gaussian of small σ, find attributes of these blobs, and then use a statistical test to distinguish local distributions of blob attributes. While their technique gives acceptable results for some examples where the blobs are relatively well defined, it is not evident that discrete well defined blobs can always be found in grey scale images. Certainly multiple scales would be needed, and spurious links between blobs may result in significant errors in computed attributes. We believe that any scheme where the first stage is intensity edge detection is going to suffer from similar problems.

Kashyap and Eom[19] characterize textures by a *long correlation* model with a small number of parameters. These parameters $[c, d, \alpha]$ are estimated by a least squares method in the frequency domain, and the existence and location of texture boundaries estimated by a maximum likelihood technique. Good experimental results are shown on a number of images composed of textures from [8]. The Kashyap–Eom scheme, indeed any scheme which characterizes textures by parameters estimated from the Fourier power spectrum, cannot be used to discriminate textures with identical Fourier power spectra. Now, some such textures can in fact be easily discriminated by humans e.g. the even-odd texture from [18] or our bright bar–dark bar texture (Fig.3).

We have chosen to approach the problem of texture segmentation by attempting to replicate human behavior, which is far superior to that of any existing machine at this task. This resolves the fundamental difficulty in texture segmentation where we wish to ignore 'inessential variation' and find the 'semantically meaningful' boundaries e.g. between grass and gravel or between leopard skin and tree leaves. What is 'inessential' and what is 'semantically meaningful' is difficult to state in mathematical terms independent of the environments typically encountered. If we define this problem to be one of finding exactly those boundaries found by human observers, we can use psychophysical experiments to provide objective criteria of success. This should be useful for most machine vision applications. Of course there are applications like textile, tree bark or aerial imagery classification where machines could potentially perform better than humans. But for the immediate future, replicating human performance would be a very useful accomplishment.

This paper is organized as follows: We begin by reviewing previous theories of human texture perception in section 2. In section 3, our computational model of human texture perception is developed. In section 4 we argue that only even-symmetric mechanisms are used in texture perception. Section 5 contains some experimental results. Texture boundaries extracted on a set of images are shown. Also a quantitative comparison with psychophysical data on the discriminability of various textures is presented. We conclude with a brief discussion in section 6.

2 Theories of human texture perception

Major theories of texture perception due to Julesz [7,17] and to Beck [2,3] 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 grey scale images (though see Voorhees & Poggio [32] for a definition of textons for grey-scale images). Experimental results critical of these theories have appeared [13,15]. An alternative approach [10,29,6] has been to exploit the linear mechanisms (psychophysically observed spatial frequency channels and neurophysiologically observed blob, bar- and edge- sensitive neurons) which have been used to explain a range of phenomena in early spatial vision. Some experiments [6,4] suggest that this approach may explain texture perception better than the more symbolic, feature based approach of Beck and Julesz. However no scheme in this framework has been fully specified, implemented and successfully tested. The crucial experimental test is the following: Does the model correctly predict the texture boundaries found preattentively by human observers, both in images of natural scenes and the synthetic stimuli from psychophysics literature? Even better, does it correctly predict the *degree* of discriminability for different texture pairs as measured by psychophysical experiments?

3 A model of human texture perception

Seeking to construct the 'simplest' such model, we examine three successively more elaborate schemes (1) a purely linear model (2) a model based on half-wave rectified responses of linear mechanisms (3) a model with half-wave rectification and nonlinear intracortical inhibition. Models (1) and (2) prove inadequate; model (3) successfully passes the experimental test. All these models share the implicit assumption that texture discrimination is based on comparing spatially averaged responses $[\iint_T R(x,y)\,dx\,dy]/[$ Area of $T]$ over textured regions T_1, T_2. R is expected to be some simple, locally computable function of the image. It can be seen [21] that this assumption is another way of expressing the insight (due to Beck and Julesz) that in preattentive texture perception, precise positional relationships between the textons are irrelevant; only texton densities matter.

To see that a theory based purely on linear mechanisms is inadequate, consider two textures T_1, T_2 which 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 value 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 [18], or indeed any discriminable texture pair with identical first order global statistics. A generalization of this observation to n^{th} order statistics and n^{th} order polynomial operators may be found in Kube [21].

Next we try a model based on the responses of linear mechanisms followed by half-wave rectification. This is perhaps the most obvious choice of nonlinearity–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
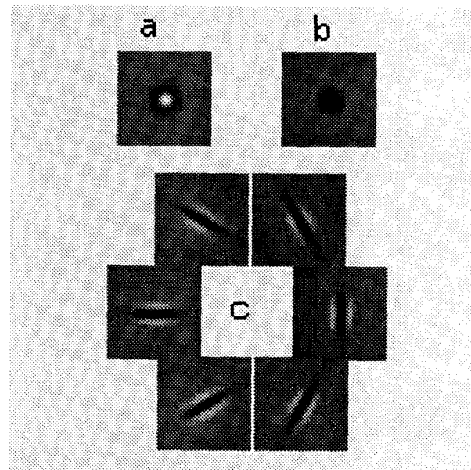

Figure 1: The point-spread functions of some of the filters used in our simulation. The filters were designed after Young [33] summing gaussian functions $G(x_0, y_0, \sigma_x, \sigma_y) = \frac{1}{2\pi\sigma_x\sigma_y}e^{-((\frac{x-x_0}{\sigma_x})^2+(\frac{y-y_0}{\sigma_y})^2)}$. (a) DOG2($\sigma$) - Linear combination of three circular concentric gaussian functions, $a \cdot G(0,0,\sigma_i,\sigma_i) + b \cdot G(0,0,\sigma,\sigma) + c \cdot G(0,0,\sigma_o,\sigma_o)$ with variance σ_i : σ : σ_o in a ratio of 0.62 : 1 : 1.6 and a : b : c in a ratio of 1 : -2 : 1. (b) DOG1(σ) - Linear combination of two circular concentric gaussian functions, $a \cdot G(0,0,\sigma_i,\sigma_i) + b \cdot G(0,0,\sigma_o,\sigma_o)$, with variance σ_i : σ : σ_o in a ratio of 0.71 : 1 : 1.14 and coefficients a : b in a ratio of 1 : -1. (c) DOOG2(σ,r,θ) - Linear combination of three offset identical gaussian functions $a \cdot G(0,y_a,\sigma_x,\sigma_y) + b \cdot G(0,y_b,\sigma_x,\sigma_y) + c \cdot G(0,y_c,\sigma_x,\sigma_y)$. Variances $\sigma_y = \sigma$, $\sigma_x = r \cdot \sigma$, offsets $y_a = -y_c = \sigma$, $y_b = 0$, coefficients a : b : c in a ratio of -1 : 2 : -1 for the filter with axis of symmetry along the x direction ($\theta = 0$). The other DOOG2() filters are obtained by rotation about the center of the middle gaussian. The scaling coefficients a_{DOG1} : a_{DOG2} : a_{DOOG2} were in a ratio of 3 : 4.15 : 2 designed to equalize the dynamic range of the respective responses.

cells are needed to represent the positive and negative parts of the response. This gives the following model: the image $I(x,y)$ is convolved with a bank of linear filters F_i followed by half-wave rectification to give a set of 'neural' responses $R_i^+(x,y)$ (the positive part of the response i.e. $\max(R_i(x,y),0)$) and $R_i^-(x,y)$(the negative part, i.e. $\max(-R_i(x,y),0)$). A texture T_1 can be preattentively discriminated from T_2 if and only if one of these 'neural' responses, say R_j^+, after spatial averaging is 'sufficiently' different for T_1 and T_2.

3.1 Choice of the filters

To specify a functional form for the filters F_i, we follow Young [33] and use radial and directional Gaussian derivatives which give excellent fits to cortical receptive field data. This choice is convenient but not critical—Gabor filters could have been used instead. The radially symmetric filter classes DOG1(σ) and DOG2(σ) (Fig 1-(a),(b)) model non-oriented simple cells. Directionally tuned filters DOOG2 (σ,r,θ) with an even-symmetric cross section perpendicular to their axes (Fig 1-(c)) model bar–sensitive
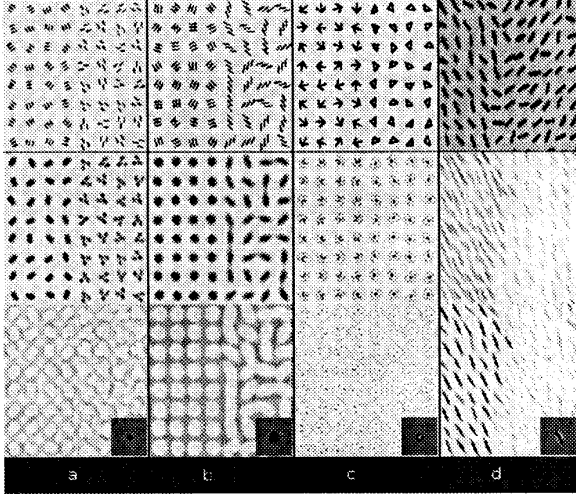
Figure 2: Some textures (top row) and half-wave-rectified response of one of the filters to each (bottom row). The point-spread function of each filter is shown at the bottom-right corner of the response image. The filter shapes are as in Figure 1 ; the frequency parameters correspond to a 4° × 4° image. 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 [15], fig. 6, pair 2.2 (top); response of a 8 c/deg DOG1 filter (bottom); $\sigma \approx 0.5 \times$(length of texel line segments). (b)Texture from [15], fig. 6, pair 2.1 (top), response of a 5 c/deg DOG1 filter (bottom); $\sigma \approx 2 \times$(width of texel line segments). (c)Arrow-Triangle texture (top), the arrow texel is obtained from the triangle by shifting one of its legs; response to a 5 c/deg DOG2 filter (bottom); $\sigma \approx 0.3 \times$(length of triangles' hypotenuse). (d)Texture from [31], fig. 4.2b (top); response to a 13 c/deg DOOG2 filter (bottom); $\sigma_y \approx$(width of bars), $\sigma_x : \sigma_y = 3$ and orientation 120°.

simple cells. In our simulations we used 6 equally spaced orientations θ and a constant aspect ratio $r = 3$. Implicit in the Gaussian derivative model is the assumption that receptive field profiles in the direction perpendicular to their axes are either odd–symmetric or even–symmetric, and not of intermediate phase. Psychophysical studies on phase discrimination [14,9] provide strong evidence for this assumption. We make the claim that only even–symmetric mechanisms (Fig. 1-(a),(b),(c)) play a role in texture perception; we ignore the DOOG1 filters of Young [33] which model odd-symmetric mechanisms which respond optimally to appropriately oriented edges. Justification for this is provided in Section 4. The σ parameter of the three filter classes used corresponds to a nominal spatial frequency in c/deg (given the viewing distance and size of image). To sample adequately the mid spatial frequency range around the peak of the luminance contrast sensitivity function, all integer values of the frequency between 3 and 14 c/deg were used. This gives 96 filters F_i which result in 192 'neural' responses R_i^+, R_i^-. We will refer to these 192 responses as 'channels'.

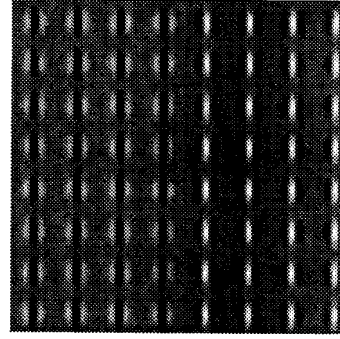Representative examples of these responses for some textures may be found in Fig. 2.

Figure 3: An easily segmentable texture pair. The micropatterns $M_1(x,y)$ on the left are bright bars DOOG2($\sigma,r,\pi/2$) from Fig.1c, and on the right $M_2(x,y)$ are dark bars $-$DOOG2($\sigma,r,\pi/2$). $M_1(x,y)$ and $M_2(x,y)$ have the same average brightness. Let the responses of these micropatterns to the filter DOOG2($\sigma,r,\pi/2$) be $R^1(x,y)$ and $R^2(x,y)$ respectively. Now $R^1(x,y) = -R^2(x,y) =$ DOOG2($\sigma,r,\pi/2$) * DOOG2($\sigma,r,\pi/2$) $= G(2\sigma,0,\sqrt{2}\sigma,\sqrt{2}r\sigma) -$ $4G(\sigma,0,\sqrt{2}\sigma,\sqrt{2}r\sigma)$ $+$ $6G(0,0,\sqrt{2}\sigma,\sqrt{2}r\sigma)$ $-$ $4G(-\sigma,0,\sqrt{2}\sigma,\sqrt{2}r\sigma) + G(-2\sigma,0,\sqrt{2}\sigma,\sqrt{2}r\sigma)$. Note that $\int\int R^1 \, dx \, dy = 0$ which means that the area under the positive part of R^1, R^{1+} is equal to the area under R^{1-}. Note also that $R^{1-} = R^{2+}$. It follows that the spatially averaged responses in the bright bar channel (positive half-wave rectification) are the same for the two micropatterns.

3.2 Half–wave rectification is insufficient

The model of texture discrimination suggested above successfully explains human performance on a number of examples. However it cannot explain the easy segmentation of the texture in Fig. 3. Consider the channel corresponding to the positive half-wave rectified response to a vertical bright bar tuned DOOG2($\sigma,r,\pi/2$) filter. One might expect this to give maximum response to the bright bars in the left region and minimum response to the dark bars on the right. Not so. As shown in Fig. 4, the spatially averaged responses are identical. The area under the response to the bright bar (Fig. 4(b)) is equal[1] to that under the response to the dark bar (Fig. 4(c)). We'll refer to the bright bar as the *optimal* stimulus for the channel (gives peak response for a given norm of the stimulus)–the dark bar is an example of a *non-optimal* stimulus which gives a *spurious* response in this channel.

This example is merely an instance of a general phenomenon with linear filters. A vertical bright bar of width σ would give responses on bar channels of other orientations and width σ's, and also on the channels corresponding to the radially symmetric DOG1 and DOG2 filters. The peak values on the non-optimum channels are smaller, but linear spatial summation of the diffuse spurious responses can give similar or even exactly equal results.

3.3 Nonlinear inhibition

A theory of texture perception must therefore consider other cortical nonlinearities—half-wave rectification alone is insufficient. In our theory we rely on intracortical inhibition. A nonlinear aggregated contrast response function (for individual neurons see [1])

---

[1]In fact, this is true for any texture pair where the constituent micropatterns M_1, M_2 have zero DC values and satisfy $M_1 = -M_2$
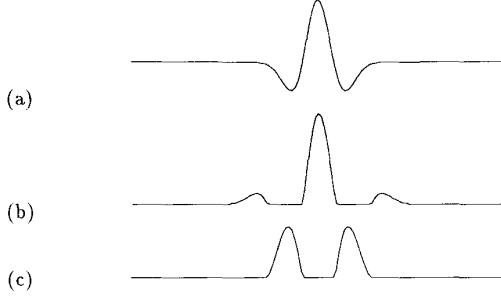
(a)

(b)

(c)

Figure 4: Cross-sections along the x-axis of the graphs of three functions: (a) M_1 (bright bar), (b) R^{1+} the response to M_1 in the bright bar channel, (c) R^{2+} the response to the dark bar M_2 in the bright bar channel (equivalently R^{1-} the response to M_1 in the dark bar channel). The areas under (b) and (c) are equal.

would be another option; the difficulty would be in coming up with a principled choice of the nonlinearity.

Inhibitory connections and interactions (some nonlinear) among the neurons in primary visual cortex V1 have been well documented [27,28,12,26]. A number of functional roles have been attributed to these, including the generation or sharpening of orientation and length selectivity.

We claim that another consequence of intracortical inhibition is the suppression of 'spurious' responses. A rigorous justification of this claim requires quantitative models of the relevant inhibitory circuits—current experimental data is inadequate to support such an exercise. We are left with the option of proposing a computational model and arguing 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—in preattentive processing there is time for just a few spikes (3) not demand unduly specific interconnection strengths between arbitrary pairs of neurons.

Our scheme is simple: Thresholds $T_i(x_0, y_0)$ for neurons belonging to channel i with retinotopic coordinates x_0, y_0 are computed

$$T_i(x_0, y_0) = \max_{j} \max_{x,y \in I_{ji}(x_0, y_0)} \alpha_{ji} R_j(x, y)$$

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, α_{ji} is a measure of the effectiveness of this inhibition. The post–inhibition response $PIR_i(x_0, y_0)$ is given by

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

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.

I_{ji} and α_{ji} are chosen by the design criterion of eliminating 'spurious' responses. As an example, consider channels i, j corresponding to the positive and negative responses of the filter $DOOG2(\sigma)$. Figs. 4(b)-(c) shows 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 by 1.25σ from it. This prompts a choice of $\alpha_{ji} = 0.65$ and

$i \downarrow \qquad j \rightarrow$	$DOG1(\sigma)$	$DOG2(\sigma)$	$DOOG2(\sigma,r,\theta)$
$DOG1(\sigma)$	0.2	0.25	0.15
$DOG2(\sigma)$	0.25	0.25	0.20
$DOOG2(\sigma,r,\theta)$	0.15	0.20	0.65

Figure 5: Inhibitory coefficients α_{ji}

$i \downarrow \qquad j \rightarrow$	$DOG1(\sigma)$	$DOG2(\sigma)$	$DOOG2(\sigma,r,\theta)$
$DOG1(\sigma)$	2σ	1.5σ	1.25σ
$DOG2(\sigma)$	2σ	1.5σ	1.25σ
$DOOG2(\sigma,r,\theta)$	2σ	1.5σ	1.25σ

Figure 6: Radii of inhibition neighbourhoods I_{ji}.

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 as specific interconnection strengths are required between arbitrary pairs of neurons. We can however exploit the known clustering: non–orientationally tuned neurons tend to occur in the V1 blobs, and neurons sharing similar orientation preference occur together. This clustering leads us to form 8 groups of channels in our framework (2 radially–symmetric + 6 oriented). I_{ji} and α_{ji} are identical for all channels i in one of these groups, these values having 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 tabled in Figures 5 and 6. No attempt has been made to optimize these numbers. We believe our model of inhibition to be too tentative a guess of the biological reality to justify such an exercise.

3.4 Computation of the texture gradient

Our model for texture discrimination is now in its final form: two textures T_1, T_2 are discriminable if their spatially averaged responses in some channel PIR_i are 'sufficiently' different. To construct a scheme for finding texture boundaries in images from the post-inhibition responses, we note the experimental findings of Nothdurft [23]. The visual system responds to a texture gradient, rather than absolute differences. 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. Biologically, the computation of the gradient of the smoothed post-inhibition response in each channel can be done 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 RFs (σ' is a measure of the size) and presumably occur in some extrastriate area. Texture boundaries are marked at local peaks of the texture gradient magnitude. An alternative way of defing the texture gradient would be as the sum of the gradients in the separate channels, instead of the maximum. We have not yet experimented with this option.

4 Odd-symmetric mechanisms are not used

Our model used only channels corresponding to even-symmetric filters. This choice was based on an interpretation of some experimental results of Rentschler, Hubner and Caelli [24] who found that textures composed of mirror-image compound Gabor signals
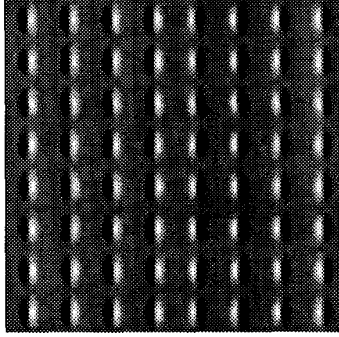
Figure 7: A texture pair composed of y-mirror-symmetric micropatterns. Segmentation is not preattentive. Compare with Figure 3 and see text.

were indistinguishable even when the individual micropatterns were easily discriminated. There was no difficulty in discriminating textures composed of non-mirror-image compound Gabor signals. A simplified version of the phenomenon can be seen by comparing Fig. 3 (easily segmentable) and Fig. 7 (not preattentively segmentable). We attribute the failure to discriminate preattentively textures of mirror-image micropatterns to a lack of utilization of channels corresponding to odd-symmetric filters.

First 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 $M_1(x) = -M_2(-x)$. Examples of y-ms pairs are Figs. 2a, 2c in Ref. [24] and the two micropatterns in our Fig. 7; Fig. 3 contains a xy-ms pair. Consider any two y-ms patterns M_1, M_2. Now, the following operations (or any composition thereof) preserve y-mirror-symmetry: (a) halfwave rectification (b) convolution with any even-symmetric filter (c) nonlinear scaling $I \mapsto g(I)$. 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 post-inhibition responses $PIR_i(M_1)$, $PIR_i(M_2)$ 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 which are a y-ms pair have identical spatial averages; from the preceding argument so must post-inhibition responses in even-symmetric channels. In other words, to segment a texture composed of M_1 from one composed of M_2 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. Now that we have identified texture pairs whose discrimination must rely on exactly one of the two symmetry classes of mechanisms, deciding which of these are used in texture perception becomes an empirical question. Clearly odd-symmetric mechanisms are not utilized; even-symmetric are.

A second argument against the use of odd-symmetric mechanisms is based on data on how our ability to discriminate textures scales with eccentricity of viewing direction (from fovea to periphery). Saarinen, Rovamo and Virsu [25] showed on a number of textures that if the stimuli were M-scaled (scaled to make them equally visible at all eccentricities, a magnification approximately
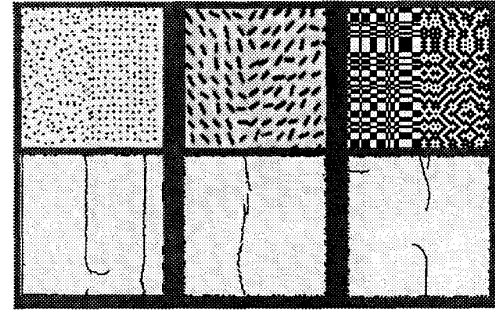
(a)-(b)-(c)
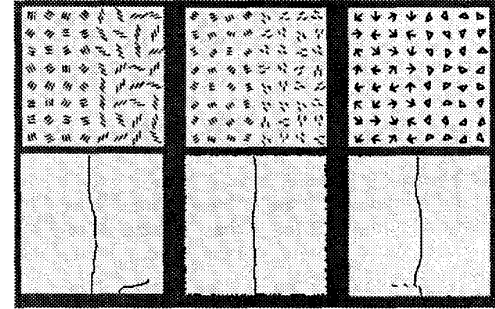
(d)-(e)-(f)

Figure 8: Texture boundaries found by our computational model on a set of images. (a) a texture where second order statistics are different in the two regions (b) a texture pair where the textons have differing orientations in the two regions (from [31]) (c) is the even-odd texture pair from [18] which can be segmented even though the left and right regions have identical third-order statistics (d) and (e) Two stimuli constructed after Gurnsey and Browse [15] who used them to criticize the original Julesz texton theory (easily segmentable in spite of identical textons in left and right regions). (f) the triangle-arrow texture pair. In our model, center–surround DOG2 filters contribute most to the segmentation and their responses decrease when the micropatterns are anisotropically stretched explaining the observations of Enns [13].

the inverse of the density of retinal ganglion cells as a function of eccentricity), then texture discrimination was equally possible at all eccentricities. Bennett and Banks [5] carried out compound grating phase discrimination experiments at a range of eccentricities. They found that the data could be explained by assuming that the sensitivity of even–symmetric mechanisms is constant, but that of the odd-symmetric mechanisms falls dramatically with eccentricity. This accounts, in their opinion, for the reduced ability to encode phase peripherally. This decline in sensitivity of the odd-symmetric mechanisms seems to suggest that they are not utilized in texture segmentation in an essential way–if they were our ability to discriminate textures should decline at a similar rate.
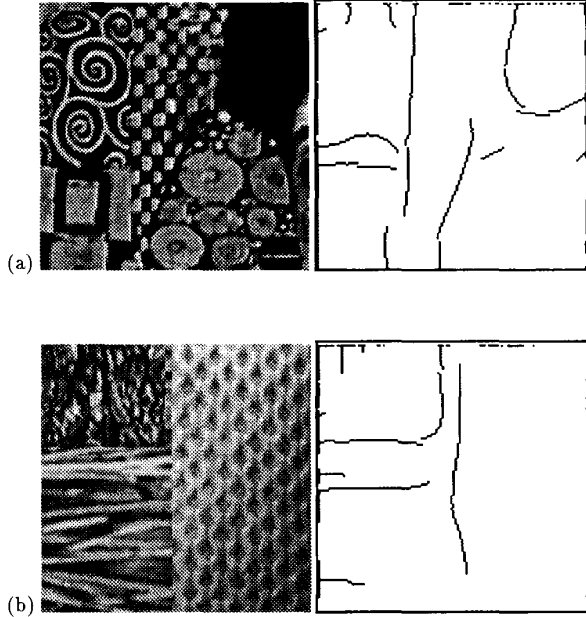
5 Experimental results

330

Figure 9: (a) A detail of the portrait of Adele Bloch-Bauer by Gustav Klimt (left) and the texture boundaries found (right). The essential boundaries of the 5 perceived groups have been detected. (b) A 'naturalistic' texture and the boundaries found. The extra 'spurious' horizontal boundary detected in the straw texture can be explained by a local texture gradient.

(Plus-ell) (Pl23-sq23)

(Tri-arr) (Plus-ti)

(Ti-ell) (RR-RL)

Figure 10: Bipartite textures with elements designed after Kröse. The texure pairs are ordered (top to bottom, left to right) from most to least discriminable.

We carried out two series of tests on our model. For all simulations the values of the parameters α_{ji} and I_{ji} were as specified in the tables of Fig. 5 and 6.

The first series of tests was to verify whether our model does indeed correctly predict the texture boundaries found by human observers. In our algorithm, texture boundaries are marked at local peaks of the texture gradient magnitude which are above some threshold. In our implementation, the Canny [11] edge–detector is used for this step, with $\sigma' = 8$ or 16 pixels in our simulations. S_i , the sampling neighborhoods, were chosen to be disks of radii σ- the size tuning parameter of channel i. Figure 8 and 9 show the texture boundaries extracted on some images.

The second series of tests was intended to compare the *degree* of texture discriminabilty predicted by our algorithm with psychophysical data due to Kröse[20] and Gurnsey and Browse[15]. Figure 10 shows 5 bipartite textures with elements constructed after [20](Section 3.2, pp.34-39), and 1 composed of R's and mirror-image R's (called RR-RL). For each of these textures, the texture gradient ($\sigma' = 12$ pixels, S_i = constant) obtained by our algorithm is computed as a function of column number (see [22]). 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 Figure 11, this data is compared with Kröse (Table 3.1, pg. 39, SOA=320) and Gurnsey & Browse (pairs 1.1, 1.2, 1.3, 3.1) mean overall discriminability. Note that the rank order of discriminability predicted by our model matches Gurnsey & Browse's ranking exactly, and that of Kröse except for the relative order-
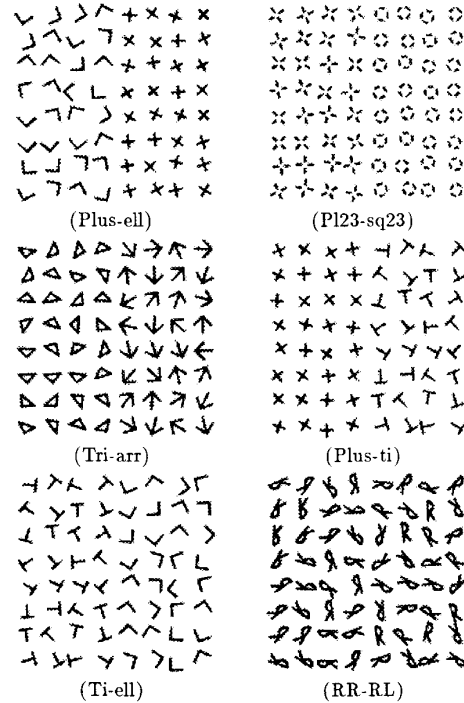
ing of the Plus-ell and Pl23-sq23 textures. The discrepancy may be explained by slight differences in the micropatterns used for building the stimuli, and the stimuli arrangement (bipartite versus rings) which can lead to foreground/background effects [15].

6 Discussion

The results in the previous section illustrate the explanatory power of our model and suggest that many of the essential aspects of texture perception have been captured in our theory. As the results in Figures 8 and 9 show, our scheme constitutes a high quality texture edge detector which works equally well on images of artificial and natural scenes. The algorithm makes use of simple, local and parallel operations which makes it potentially realtime.

Our framework can be extended to include the detection of brightness/color step edges. For that one adds channels corresponding to *odd–symmetric* directional Gaussian derivatives, which are the filters used in the Canny edge detector. Nonlinear inhibition can be used to help select the scale of the edge, not unlike the idea of non-maximum suppression in the Canny scheme. We will be reporting on this unified boundary detection scheme in a forthcoming paper.

Acknowledgements

Texture pair	Discriminability		
	Kröse	Gurnsey & Browse	Malik & Perona
Plus-ell	68.6	0.736	232
Pl23-sq23	88.1	n.a.	201
Tri-arr	52.3	0.4–0.55	154
Plus-ti	37.6	0.496	136
Ti-ell	30.6	0.421	97
RR-RL	n.a.	n.a.	50

Figure 11: A comparison of the performance of our texture segmentation algorithm with psychophysical data from Kröse [20] and Gurnsey and Browse [15]. The value of the texture gradient at column 64 in our 128x128 images is taken to be our model's discriminability prediction. Because of differences in the scales used, the three columns should be compared only by the rank ordering of discriminability. The Kröse data corresponds to δ_{tb} (Table 3.1, pg. 39, SOA=320) and Gurnsey & Browse data to mean overall discriminability (pairs 1.1, 1.2, 1.3, 3.1 in their paper) averaged over foreground/background and different stimulus durations. Note that the rank order of discriminability predicted by our model matches Gurnsey & Browse's ranking exactly, and that of Kröse except for the relative ordering of the Plus-ell and Pl23-sq23 textures. The RR-RL texture pair is not segmentable, as a lateral peak of the gradient exceeds the central peak.

References

[1] D. Albrecht and D. Hamilton. Striate cortex of monkey and cat: contrast response function. *Journal of Neurophysiology*, 48:217–237, 1982.

[2] J. Beck. *Organization and representation in perception*, chapter Textural segmentation. Erlbaum, Hillsdale, NJ, 1982.

[3] J. Beck, K. Prazdny, and A. Rosenfeld. *Human and Machine Vision*, pages 1–38. Academic Press, 1983.

[4] J. Beck, A. Sutter, and R. Ivry. Spatial frequency channels and perceptual grouping in texture segmentation. *Computer Vision, Graphics and Image Processing*, :299–325, 1987.

[5] P. Bennett and M. S. Banks. Sensitivity loss in odd-symmetric mechanisms and phase anomalies in peripheral vision. *Nature*, 326:873–876, 1987.

[6] J. Bergen and E. Adelson. Early vision and texture perception. *Nature*, 333:363–364, May 1988.

[7] J. Bergen and B. Julesz. Rapid discrimination of visual patterns. *IEEE Transactions on Systems, Man, and Cybernetics*, 13(5), 1983.

[8] P. Brodatz. *Textures*. Dover, 1966.

[9] D. Burr, C. Morrone, and D. Spinelli. Evidence of edge and bar detectors in human vision. *Vision Research*, in press.

[10] T. Caelli. Three processing characteristics of visual texture segmentation. *Spatial Vision*, 1(1):19–30, 1985.

[11] J. Canny. *Finding edges and lines in images*. Technical Report 720, MIT, Artificial Intelligence Laboratory, 1983.

[12] K. De Valois and R. Tootell. Spatial-frequency-specific inhibition in cat striate cortex cells. *Journal of Physiology*, 336:359–376, 1983.

[13] J. Enns. Seeing textons in context. *Perception and Psychophysics*, 39(2):143–147, 1986.

[14] D. Field and J. Nachmias. Phase reversal discrimination. *Journal of Vision Research*, 24:333–340, 1984.

[15] R. Gurnsey and R. Browse. Micropattern properties and presentation conditions influencing visual texture discrimination. *Perception and Psychophysics*, 41(3):239–252, 1987.

[16] R. Haralick. Statistical and structural approaches to texture. *Proceedings of the IEEE*, 67(5):786–804, 1979.

[17] B. Julesz. Texton gradients: the texton theory revisited. *Biological Cybernetics*, 54:245–251, 1986.

[18] B. Julesz, E.N.Gilbert, and J.D.Victor. Visual discrimination of textures with identical third order statistics. *Biological Cybernetics*, 31:137–140, 1978.

[19] R. L. Kashyap and K. Eom. Texture boundary detection based on the long correlation model. *IEEE transactions on Pattern Analysis and Machine Intelligence*, 11:58–67, 1989.

[20] B. J. A. Kröse. *A description of visual structure*. PhD thesis, University of Delft, 1986.

[21] P. R. Kube. *On Image Texture*. PhD thesis, University of California at Berkeley, 1988.

[22] J. Malik and P. Perona. *A computational model of texture perception*. Technical Report UCB/CSD 89/491, Computer Science division (EECS), U.C.Berkeley, February 1989.

[23] H. C. Nothdurft. Sensitivity for structure gradient for texture discrimination tasks. *Vision Research*, 25:1957–1968, 1985.

[24] I. Rentschler, M. Hubner, and T. Caelli. On the discrimination of compound gabor signals and textures. *Vision Research*, 28(2):279–291, 1988.

[25] J. Saarinen, J. Rovamo, and V. Virsu. Texture discrimination at different eccentricities. *Journal of the Optical Society of America - A*, 4(8):1699–1703, 1987.

[26] A. M. Sillito and P. C. Murphy. *Neurotransmitters and cortical function: from molecules to mind*, chapter 11. Plenum Press, 1988.

[27] D. Tolhurst. Adaptation to square wave gratings: inhibition between spatial frequency channels in the human visual system. *Journal of Physiology*, 226:231–248, 1972.

[28] K. Toyama, M. Kimura, and K. Tanaka. Organization of cat visual cortex as investigated by cross–correlation techniques. *Journal of Neurophysiology*, 46:202–214, 1981.

[29] M. Turner. Texture discrimination by gabor functions. *Biological Cybernetics*, 55:71–82, 1986.

[30] L. Van Gool, P. Dewaele, and A. Oosterlinck. Texture analysis anno 1983. *Computer Vision, Graphics and Image Processing*, 29:336–357, 1985.

[31] H. Voorhees. *Finding texture boundaries in images*. Technical Report 968, M.I.T. Artificial Intelligence Laboratory, 1987.

[32] H. Voorhees and T. Poggio. Computing texture boundaries from images. *Nature*, 333:364–367, 1988.

[33] R. Young. *The gaussian derivative theory of spatial vision: Analysis of cortical cell receptive field line-weighting profiles*. Technical Report GMR-4920, General Motors Research, 1985.