A computational model of texture segmentation

Jitendra Malik and Pietro Perona
Department of Electrical Engineering and Computer Science
University of California at Berkeley, CA 94720

Abstract

We present a computational model of human texture segmentation and argue for its utility in machine vision. Major theories due to Julesz and Beck attribute preattentive texture segmentation to differences in first-order statistics of stimulus features such as orientation, size and brightness of constituent elements. An alternative approach seeks to exploit psychophysically observed spatial frequency channels and neurophysiologically observed blob, bar- and edge-sensitive mechanisms, and perform simple computations on the outputs of these to find texture boundaries. Previous models in this framework have been incompletely specified; our model is precisely stated and applicable to arbitrary grey scale textures. We claim that the responses of two types of mechanisms are necessary and sufficient: (a) center-surround mechanisms of various widths, and (b) oriented mechanisms of various widths and orientations which are even-symmetric about their axes. Simulation data on a number of texture pairs is presented.

1 Introduction

The two primary problems in texture perception-texture classification and texture segmentation have received considerable attention in both human and machine vision. For surveys of the research in machine vision we refer the reader to Haralick [17], Van Gool, Dewaele and Oosterlinck [27] and Kube [21]. In human vision, the leading theories of texture perception are due to Beck and Julesz [7,19,2,3]. They attribute preattentive texture segmentation to differences in first-order statistics of stimulus features such as orientation, size and brightness of constituent elements. Additionally Julesz's texton theory [7,19] considers such features as terminators and intersections, and Beck[2,3] attributes a major role to emergent structures from grouping operations. Experimental results critical of these theories have appeared [13,16,6,4].

Research in texture perception has suffered from a major difficulty-vision researchers seem unable to agree on a mathematically precise definition of texture. This is obviously troubling for machine vision-if one does not know what is to be the input and output of an algorithm, how does one decide that the algorithm is effective? In section 2, we argue that machine texture segmentation should replicate human bahavior; one can then use psychophysical data to provide the input/output characterization. The rest of the paper is devoted to developing a computational model of human texture segmentation which is, by this definition, automatically a solution to the corresponding machine vision problem. In section 3 we review psychophysical and electrophysiological evidence on phase-sensitive mechanisms in primary visual cortex V1. In section 4, we present our model of texture segmentation which is based on finding differences in the pooled responses of these mechanisms. The important question then is one of precisely specifying the class of detectors used in human texture perception. We interpret the results of a set of psychophysical experiments on phase discrimination to answer this question. We claim that the responses of two types of mechanisms are necessary and sufficient: (a) center-surround mechanisms of various widths, and (b) oriented mechanisms of various widths and orientations which are even-symmetric about their axes. In section 5, we present arguments which suggest that odd-symmetric mechanisms are *not* used in human texture perception. Simulation data on a number of texture pairs is presented in Section 6.

2 Why machines should imitate human texture perception

Both texture recognition and segmentation have been hard problems in machine vision for a long time. One major problem has been the lack of an adequate definition of texture and objective criteria of success at a task involving texture perception. This is not the case for many other machine vision tasks. To take two examples, the stereo correspondence problem and the 3-D object recognition problem can be defined mathematically without reference to human abilities and one can certainly imagine machine vision systems which outperform humans. In texture segmentation the problem is the following-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 and the tasks that need to be performed. Consider one possible approach in terms of a physical variable we intuitively think of as related to texture: spatial frequency measured in cycles/degree subtended at the retina. We might try to define a particular range to correspond to texture. However we realize that a bird like a falcon or an eagle needs to be (and is !) much more sensitive to higher spatial frequencies than a human. What is inessential variation to a human is semantically meaningful for the bird. A specification of inessential variation by an absolute frequency range is clearly inappropriate.

We choose to define the problem of texture segmentation by machine to be the task of replicating human behavior at this task: find exactly those boundaries preattentively extracted by human observers. This should be useful for most machine vision applications: autonomous navigation by approximately human size robots in natural environments is an obvious example. We realize that there are applications like textile or tree bark classification where machines could potentially perform better than humans with appropriate definitions of 'inessential' or 'semanti-

22ACSSC-12/88/0490 $1.00 © 1988 MAPLE PRESS

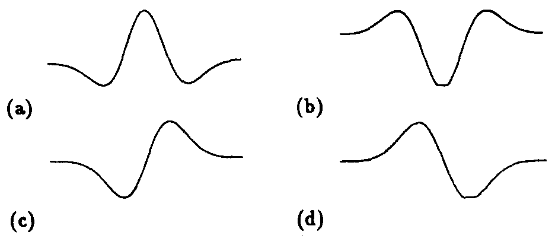

Figure 1: Four receptive field profiles. Even-symmetric (top), and odd-symmetric (bottom).

cally meaningful' variation. However replicating human behavior seems to us to be the best way of coming up with a task independent definition. Note that we have avoided giving a definition of texture; we only defined texture boundaries. We offer a suggestion due to Kube [21]: Texture is spatial variation in image irradiance that occurs within the segments of a desired image segmentation.

3 Even–symmetric and odd–symmetric mechanisms

The idea that the early visual system contains a range of channels sensitive to different spatial frequencies is due to Campbell and Robson [10] and has received considerable support from psychophysical and neurophysiological data. While some earlier studies were suggestive, phase discrimination experiments by Field and Nachmias [15] provide strong evidence for a further elaboration: there are in fact 4 distinct classes of detectors whose receptive field profiles are shown in Figure 1. Bar-selective detectors are even–symmetric (a) and (b) and respond strongly to appropriately placed bright and dark bars respectively. Edge-selective detectors (c) and (d) are odd–symmetric and respond preferentially to left and right step edges, respectively. Note that the names bar and edge detectors are primarily for mnemonic value and each mechanism may respond to other stimuli; also there may be more secondary lobes than shown in the figure. Field and Nachmias presented observers with two compound gratings composed of a fundamental (F) and second harmonic (2F) added in two phases differing by 180 degrees. Discrimination thresholds for several 180 phase shifts (for example, 0-180, 45-225, 90-270) were measured. They explained their results by a model which uses changes in the responses of even-symmetric mechanisms centered on the peaks and troughs of F and odd-symmetric mechanisms centered on the zero-crossings of F. Burr, Morrone and Spinelli [8] studied compund grating phase discrimination using stimuli composed of 256 harmonic cosine components, smoothly filtered in amplitude. They exploited the facilitation phenomenon and measured thresholds for discriminating compound gratings from their negatives (180 phase shift) in the presence of a pedestal and studied how these varied with pedestal contrast. Again their experimental results suggested strongly the presence of these 4 mechanisms (and not ones of intermediate phase). It seems natural to connect these 4 mechanisms to electrophysiologically observed simple cell receptive field profiles [18]. For computational purposes, one can model the even-symmetric mechanisms as linear convolution with a RF of type 1(a) followed by half-wave rectification: the positive part of the response gives the bright bar mechanism and the negative part of the response gives the dark bar mechanism. The half-wave rectification models the biological fact that cortical cells have low maintained discharge rates and are unable to respond with a decrease in firing rate as required by a negative response. Similar remarks apply for the odd-symmetric mechanisms.

4 A model of human texture segmentation

Several researchers [6,14,9,4,26] have suggested a role for spatial frequency channels, neurophysiological blob/bar/edge detectors, and pooled responses in texture perception. However, it is fair to say that a model precise enough to be implemented and tested (and perhaps falsified !) on grey scale textures has not yet been given. We offer such a 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 $R_i^+(x,y)$ and $R_i^-(x,y)$. 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 pooling is 'sufficiently' different for T_1 and T_2. Spatial pooling here means the computation of the average response over each textured region. R_i^+ and R_i^- should be viewed as convenient approximations of the true cortical neural responses $g(R_i^+)$ and $g(R_i^-)$ where g is a nonlinear function of the type observed electrophysiologically [1]. Two noteworthy points here: (a)A nonlinear stage is essential: spatially pooled responses of linear filters cannot discriminate between textures with identical second-order statistics and hence identical power spectra. And we know that humans can discriminate some textures with identical second-order statistics preattentively [20]. A generalization of this observation to n^{th} order statistics may be found in Kube [21]. (b) Half-wave rectification is a natural nonlinearity as stated in the previous paragraph, and it is necessary. Full wave rectification [6] will not permit us to discriminate a texture consisting of bright bars on a grey background from a texture consisting of identical size, contrast dark bars on a grey background if one uses filters F_i with zero-mean response profiles.

To completely specify the model, all that remains is the choice of filters F_i. Two classes of filters seem natural choices:(1) radially symmetric center–surround filters of the difference of gaussian type (Figure 2.a-b), and (2) directionally tuned filters of differing orientation and widths with an even-symmetric cross section perpendicular to their axes (Figure 2c). The first kind are suggested by our ability to segment texture pairs consisting of bars of differing width and orientation. The second kind are needed to discriminate texture pairs composed of micropatterns containing the same oriented line segments, such as triangles and K's, or X's and L's. Later in this paper we present simulation results, including examples from Gurnsey and Browse [16] which were sources of great difficulty for Julesz's texton theory. For now, let us examine whether orientationally-tuned odd–symmetric mechanisms (as in 1.c and 1.d) are needed or used. We suggest not, and present two arguments for this claim in the next section

5 Odd–symmetric mechanisms not used

The first argument is based on some experimental results of Rentschler, Hubner and Caelli [22] on the discrimination of textures composed of micropatterns which were compound Gabor signals. Textures composed of mirror-image compound Gabor

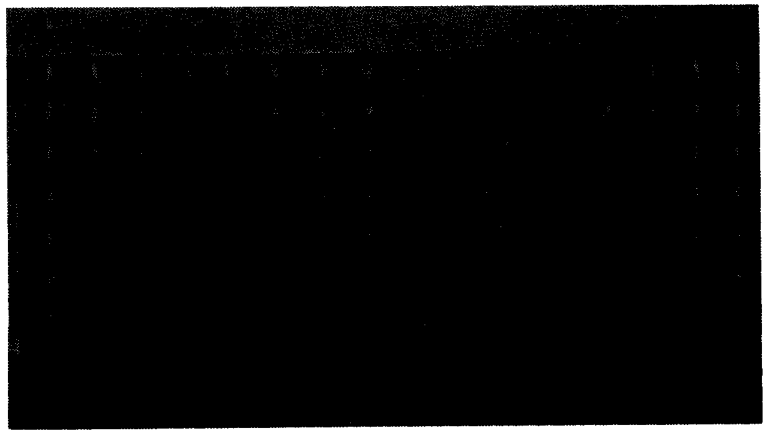

Figure 3: Two texture pairs composed of odd-symmetric (left) and even-symmetric (right) micropatterns. The micropatterns in the left texture pair are y-mirror-symmetric; the micropatterns in the right texture pair are xy-mirror-symmetric. Segmentation is preattentive for the right texture pair but not for the left one.

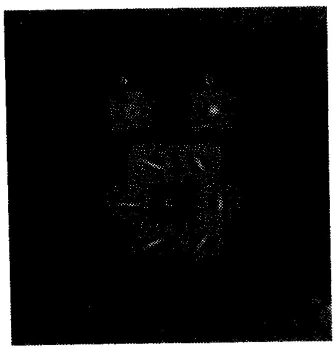

Figure 2: The point-spread functions of some of the filters used in our simulation. The filters were designed after Ref. 24 summing gaussian functions $G(x_0, y_0, \sigma_x, \sigma_y) = \frac{1}{2\pi\sigma_x\sigma_y} e^{-((\frac{x-x_0}{\sigma_x})^2+(\frac{y-y_0}{\sigma_y})^2)}$. (a) DOG2 - Linear combination of three circular concentric gaussian functions, $a \cdot G(0,0,\sigma_a,\sigma_a) + b \cdot G(0,0,\sigma_b,\sigma_b) + c \cdot G(0,0,\sigma_c,\sigma_c)$ with variance $\sigma_a : \sigma_b : \sigma_c$ in a ratio of $0.62 : 1 : 1.6$ and $a : b : c$: in a ratio of $1 : -2 : 1$. (b) DOG1 - Linear combination of two circular concentric gaussian functions, $a \cdot G(0,0,\sigma_a,\sigma_a) + b \cdot G(0,0,\sigma_b,\sigma_b)$, with variance $\sigma_a : \sigma_b$ in a ratio of $1 : 1.6$ and coefficients $a : b$ in a ratio of $1 : -1$. (c) DOOG2 - Linear combination of three offset identical gaussian functions $a \cdot G(0, y_a, \sigma_x, \sigma_y) + b \cdot G(0, y_b, \sigma_x, \sigma_y) + c \cdot G(0, y_c, \sigma_x, \sigma_y)$ The offsets are $y_a = -y_c = \sigma_y$, $y_b = 0$; the coefficients $a : b : c$ are in a ratio of $-1 : 2 : -1$ for the filter with axis of symmetry along the x direction. The other filters of the class are obtained by rotation about the centre of the middle gaussian. Six different orientations were used in all experiments. σ_y was varied between 1 and 10 pixels , and the ratio $\sigma_x : \sigma_y$ was varied between 2 and 4.

signals were found indistinguishable even when the individual micropatterns were easily discriminated. However, there was no difficulty in discriminating textures composed of non-mirror-image compound Gabor signals. A simplified version of the phenomenon (using differences of offset Gaussians [28] to construct simple odd and even micropatterns) is seen in figure 3 where the right texture is easily segmented and the left one is not. While they interpreted their data differently, we attribute the failure to discriminate textures of mirror-image compound Gabor signals to a lack of utilization of 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 Figures 2a, 2c in Ref. 19, and the two micropatterns in our Figure 3(left); Figure 3(right) contains a xy-ms pair. Consider any two y-ms patterns M_1, M_2. Clearly the means of their positive parts M_1^+, M_2^+ are identical; so are the means of their negative parts M_1^-, M_2^-. Pooled values after (or before!) halfwave rectification cannot distinguish them. Since y-mirror-symmetry is invariant under scaling with any (possibly nonlinear) function and under convolution with even-symmetric kernels, the two patterns are not distinguishable using pooled responses R^+, R^- of even filters. As y-mirror-symmetry is not invariant under convolution with odd-symmetric kernels, pooled responses of such filters (possibly after a nonlinearity) could be used for discrimination. In other words, to segment a texture composed of M_1 from one composed of M_2 using pooled responses, we *must* rely on odd-symmetric mechanisms. Interestingly, for a xy-ms pair, the situation is reversed: only even-symmetric mechanisms are useful. (To establish this, note that convolving a xy-ms pair with an odd kernel 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.

The second argument 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 [24] 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 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. They also used their model to account for Rentschler and Treutwein's [23] data on phase discrimination in the periphery. 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.

6 Simulation Results

Our choice of frequency and orientation tuning parameters of these filters was guided by psychophysics (contrast sensitivity functions) and electrophysiological data on macaque V1 neurons from DeValois, Albrecht and Thorell [12]. For the shape of these filters we used radial or directional gaussian derivatives as suggested by Young [28] which have been found to give excellent fits to cortical RF data. (This choice is not critical–we could have used Gabor filters instead.) We tested our model on a large number of texture pairs culled from the extensive literature on texture segmentation. For each discriminable pair, we were able to find one or more discriminating filters in the class described. Figure 4 displays some examples. Note that the parameters of these filters correspond plausibly to those of neurons in V1.

While it seems to have greater explanatory power than the Beck and Julesz models, our model should be viewed primarily as a first-order approximation. For example, we have ignored nonlinearities arising from the cross-channel inhibition in V1 which has been documented both physiologically and psychophysically [25,11]. Also in this paper we have not precisely specified how texture boundaries are computed from the neural response profiles. These are topics of ongoing investigation.

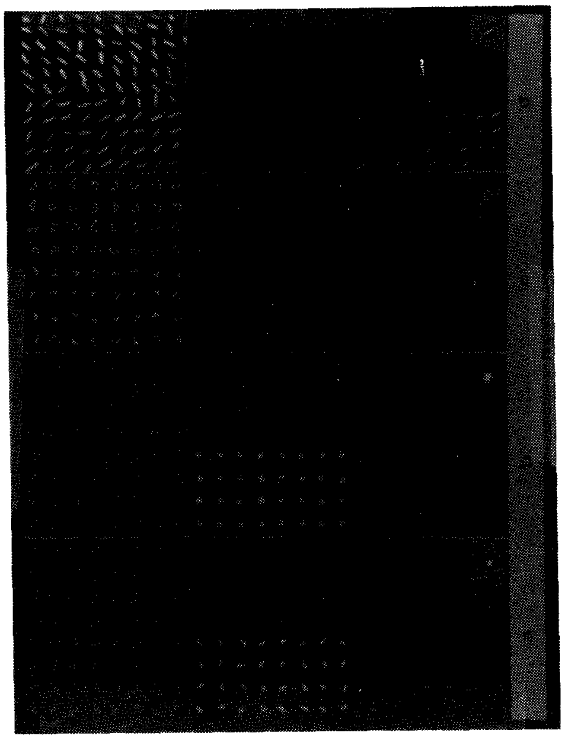

Figure 4: Some textures (top row) and half-wave-rectified response of one of the filters to each (bottom row). The point-spread function of each filter is shown at the bottom-right corner of the response image. The filter shapes are as in Figure 2; the frequency parameters correspond to a $4° \times 4°$ 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 Ref. 6, 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). Similar responses for 4–9 c/deg DOG1, 3–6 and 10–13 DOOG2 filters. (b)Texture from Ref. 6, fig. 6, pair 2.1 (top), response of a 5 c/deg DOG1 filter (bottom); $\sigma \approx 2\times$(width of texel line segments). Similar responses for 4–8 c/deg DOG1 filters. (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). Similar responses for 4–9 c/deg DOG2 filters. (d)Texture from Ref. 25, 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°. Similar responses for 4–15 c/deg DOOG2 120° filters.

Acknowledgements

We thank Paul Kube and Martin Banks for many useful discussions. This research was funded in part by an IBM Faculty Development Award and by SRC.

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

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

[10] F. Campbell and J. Robson. Application of fourier analysis to the visibility of gratings. *Journal of Physiology*, 197:551–566, 1968.

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

[12] R. De Valois, D. Albrecht, and L. Thorrell. Spatial frequency selectivity of cells in macaque visual cortex. *Vision Research*, 22:545–559, 1982.

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

[14] O. D. Faugeras. Texture analysis and classification using a human visual model. In *Proceedings, 4th International Conference on Pattern Recognition*, pages 549–552, 1978.

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

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

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

[18] D. Hubel and T. Wiesel. Functional structure of macaque visual cortex. In *Proceedings of the Royal Society - London*, pages 1–59, 1977.

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

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

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

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

[23] I. Rentschler and B. Treutwein. Loss of spatial phase relationships in extrafoveal vision. *Nature*, 313:308–310, 1985.

[24] 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.

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

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

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

[28] 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.