THREE PROCESSING CHARACTERISTICS OF TEXTURE DISCRIMINATION

Terry M. Caelli

Department of Psychology, The University of Alberta

Edmonton, Alberta T6G 2E9

Abstract

In this paper we examine the idea that texture segmentation comes about by the differential outputs of detectors (non-linear associative filters) computed at each resolvable position on the textured surface. Further, we consider some of the conditions under which "primary" detector outputs are dynamically compared and associated to develop into a smaller set of "texton" profiles which capture the predominant differentiating features of the texture regions. Comparisons to human psychophysical results are made.

KEY WORDS: texture segmentation, associative networks, orientation detectors, adaptability

1. Introduction.

For a biological visual system endowed with a multitude of cells which apparently act as feature extractors or filters, it seems reasonable to presume that visual texture segmentation may come about by the differential responses of such detectors over the textured region. This proposal has received experimental and mathematical attention over the past decade with one-dimensional grey-scaled textures (Richards, 1979; Harvey & Gervais, 1978) and two-dimensional textures (Caelli & Julesz, 1978; Caelli, 1982, 1985). However, only until recently has a full computational model been proposed which produces segmentation as a function of such "texton" (Julesz, 1981) outputs, and this paper extends the above analyses in a number of ways (Caelli, 1985).

Here texture segmentation is viewed as having three component processes: (1) spatial decomposition, (2) dynamical associative processing, and (3), classification of textured regions. The specific aims of this model are to enable segmentation when the textures consist of sparse micropatterns; to create networks which will extract, or adapt to, the predominant features of the texture; and to use a classification procedure which is adaptive to the outputs of such detectors.

2. The Model.

2.1 Level I processing: Spatial decomposition and activity profiles.

The initial process of texture segmentation is envisaged to involve the registration of the input (foveal) texture through the parallel outputs of many detectors whose responses are determined by some non-linear transformation of their cross correlation with the input. Assuming a relatively fixed "retinal pre-processor", having opponent center-surround receptive fields, the primary information to be processed must have differential, or band-pass, components emphasized. Further to this, we assume the existence of a relatively fixed primary projection area where such image derivative information is further classified (encoded) by cortical edge and bar detectors whose outputs are a non-linear function of the cross-correlation of the detector's profile with the input image. That is, we assume that the response $R_i(x,y)$ of a detector d_i at retinotopic position (x,y) is determined by:

$$\sum_{\alpha,\beta} d_i(\alpha,\beta) = \text{const.} \quad (1)$$

and $R_i(x,y) = \text{const} + Y\psi\{d_i o I\}$, $Y \equiv$ constant. (2)

o denoting cross-correlation between the detector and image (I)

$$d_i o I(x,y) = \sum_{\alpha,\beta} d_i(\alpha,\beta) I(x+\alpha, y+\beta), \quad (3)$$

and

$$-1 \leq \psi_\delta(z) = \frac{1-e^{-\delta z}}{1+e^{-\delta z}} \leq +1, \quad \delta \equiv \text{constant.} \quad (4)$$

In our simulations we have used

$$R_i(x,y) = 128 + 127\ \psi\delta\{d_i o I\}, \quad \delta = 0.03, \quad (5)$$

to fit in with an 8-bit response range. The non-linear transducer enables one to move smoothly from square wave ("ideal" edge and bar) to gaussian modulated sinusoid (Daugman, 1983) representations for edge and bar, or orientation detectors, via δ in (4). Orientations and sizes of the detectors were chosen to fit with a large number of experimental results on human texture discrimination showing the inability of observers to resolve image orientations to better than ±5° (Caelli, 1982; Beck, 1983). With evidence that such receptive, or "perceptive" fields are limited in size to ±1/8 octave, or 1 1/2 cycles to 1/e decay of a gaussian aperture, we have generated 24 fundamental orientation detectors over 7x7 pixel kernels (relative to 128x128 pixel textures) satisfying these profile constraints for both edge (odd) and bar (even) detectors.

We secondly assume that the response profile for each detector is 'rectified' into an "activity" profile.

$$A_i(x,y) = |R_i(x,y) - \text{const}|, \quad \text{const} = 128. \tag{6}$$

2.2 Level II procesing: Adaptive control and selection of critical features.

In contrast to representing texture codes by detectors defined at different size scales, in gaussian pyramids, etc. (see Burt & Adelson, 1983), which would be capable of responding to texture regions in areas greater than the actual micropattern size--the approach adopted here remains at the resolution of the basic texture--though this is not a necessary restriction. Further, the process of texture region "filling-in" (impletion) is seen as a dynamic process involving the iterative activity of activated detectors in terms of how their responses may spread over contiguous regions in a summative (averaging) fashion. This is analogous to relaxation in image processing (Hummel & Zucker, 1983) where the strength of a given spatial response is reinforced or inhibited as a function of neighbouring collaborative or opposite evidence. In particular, it is assumed that the activity of a given detector d_i determined by (6) is updated dynamically by the following (associative) "texture processing equation":

$$A_i^{t+1}(x,y) = \frac{1}{\alpha\beta} \sum_{\alpha} \sum_{\beta} A^t\left(x+\alpha-\frac{\alpha}{2}, y+\beta-\frac{\beta}{2}\right) + \sum_{\substack{j=1 \\ i\neq j}}^{n} w_{ij}^t A_j^t(x,y) \tag{7}$$

where (α,β) corresponds to the "region of influence" at each iteration. W^t_{ij} is the coupling, or associativity, between two activity profiles which can either be fixed or adaptive. For the fixed case we have used the well-known "mass-action" formulation for detector coupling (Grossberg, 1982), where we have set:

$$\forall t,\ w_{ij}^t = \begin{cases} k \ldots (i,j) \text{ being (edge, bar) pairs at the same orientation} \\ -\left(\frac{1}{n-2}\right) k \ \ldots . \text{otherwise.} \end{cases} \tag{8}$$

for n detectors and k being the coupling strength such that

$$\forall i,\ \sum_{\substack{j=i \\ i\neq j}}^{n} w_{ij}^t = 0. \tag{9}$$

In general, it seems unlikely that the (neural) connectivity between such detector planes can be defined by a single stationary matrix W_{ij} over space and time. Like Hebb (1949) and Fukushima (1984), we assume that the process of perceptual learning and adaptation involves the dynamic updating of W_{ij} as a function of the detector's response strengths and correlations. For these reasons, our interests are also focused upon investigating formulations for W^t_{ij} such as:

$$w_{ij}^t = r_{ij}^p \{a_{ij} A_j^t(x,y) + b_{ij}\} \tag{10}$$

where a_{ij} and b_{ij} correspond to slope and intercept regression coefficients of A_j on A_i, p to the degree to which this correlated information is combined with the response of A_i to result in "new" detector profiles. This dynamical system converges to strong "attractor" detectors which (as will be shown) have receptive fields related to the first few eigenvalues of the coupling matrix. Using (10) in (7) requires normalization as:

$$z_i^{t+1}(x,y) = \frac{1}{\left[1+\sum_{\substack{j=1 \\ i\neq j}}^{m} r_{ij}^p\right]} \left[\frac{1}{\alpha\beta} \sum_{\alpha\beta} z_i^t(x+\alpha, y+\beta) + \sum_{\substack{j=1 \\ i\neq j}}^{m} r_{ij}^p [a_{ij}^t z_j^t(x,y) + b_{ij}^t] \right] \tag{11}$$

for $0 \leq p$ and even. The first component:

$$\frac{1}{\alpha\beta} \sum_{\alpha\beta} z_i^t(x+\alpha, y+\beta)$$

being an averaging process (moving spatial window), is clearly "local" and restricted to a given detector plane. That is, activity within a given plane spreads as a function of the neighbouring activity of the same detector type and converges to a mean level of activity. Secondly, this activity is reinforced as a function of the degree to which other detectors exhibit similar graded responses over the full texture regions--a <u>global cooperative component</u>--represented by the last component in (11): "synergesis".

This has the effect of combining <u>correlated</u> detector responses and converging to common ("attractor")

profiles, so reducing the number of different detectors. Again, it should be noted that the solutions are critically dependent on the input signal. In the case of texture segmentation, where broad spatial regions have to be "labeled", inhibitory forms of W_{ij} seem inappropriate as they differentiate the detector responses in further ways. This, in turn, would not produce the percept of contiguous spatial regions, and would be more useful in pattern recognition where it is precisely these differentiated dimensions which are needed.

Finally, we consider the network to "complete" its activity when it reaches near equilibrium state; as measured by:

$$\frac{1}{n\alpha\beta} \sum_{x,y}^{\alpha,\beta} \left[\frac{A_i^{t+1}(x,y) - A_i^t(x,y)}{A_i^t(x,y)} \right] \leq \delta, \quad (12)$$

δ being near zero (in our case δ=0.02). Here n corresponds to the number of detectors and (α,β) to the image size.

It should be noted that formulation (11) is an example of an associative network whose coupling undergoes adaptation, and, if we consider the problem of texture segmentation as primarily involving the extraction of the main dimensions for segmentation, then it is the eigenvalues of W_{ij} which are critical. Further, we could also claim, like Kohonen (1977) and Anderson, Silverstein, Ritz and Jones (1977) that W_{ij}^t--the network associativity at time t is the primary attribute of the model rather than the detector states, per se. However, the author feels these claims to be too strong since both $\{A_i\}$ and $\{W_{ij}\}$ are mutually dependent. However, in the simulations to be reported we shall observe the behaviour of the eigenvalues of W_{ij} to investigate how these adaptive processes are changing the dimensionality of the problem.

2.3 Level III processing: Region classification and decision criteria.

Since textured regions are proposed to appear as a function of position response differences in "feature space", the appropriate classification process seems to be the minimum distance classifier (MDC, Ahmed & Roa, 1975). This method determines whether a pixel falls into one of two textured regions as a function of whether it is closer to the centroid of the texture or not. The MDC determines the discriminant (function) hyperplane which constitutes the locus of points equidistant betwen both centroids, and of the form:

$$g(a_1,\ldots,a_n) = \sum_{j=1}^{n} [\bar{a}_{1j} - \bar{a}_{2j}]a_j - \frac{1}{2}\left[\sum_{j=1}^{n} \bar{a}_{1j}^2 - \sum_{j=1}^{n} \bar{a}_{2j}^2\right] \quad (13)$$

where $(a_1..a_n)$ correspond to the feature dimensions--in this case detector outputs. $\bar{a}_{ij}$ corresponds to the mean value for group (texture) i on feature j, while a_j corresponds to a given input texture pixel feature weights. The pixel is classified as a function of the sign of $g(a_1..a_n)$.

To introduce a degree of "fuzziness", or "segmentation strength", into this procedure, it would be adequate to use the distance between the means over the standard deviation of both sets (or average t statistic):

$$\bar{t} = \frac{1}{n} \sum_{i=1}^{n} t_i, \text{ for } t_i = \frac{\bar{a}_{1i} - \bar{a}_{2i}}{\sqrt{S_{1i}^2/n_1 + S_{2i}^2/n_2}} \quad (14)$$

where n_1, n_2 correspond to the number of pixels in each textured region, s_i^2 to the appropriate variance statistic.

This function not only indicates that adding common features to the textures would decrease perceptual segmentation, but would also decrease if more variability in detector outputs was observed over either, or both, regions.

3. Simulations and Conclcusions.

We first summarize the main properties of the model:

(T_1) Detector activity is determined by the rectified response profiles as a result of detector cross-correlation with the incoming texture, according to (1)-(7).

(T_2) The activity of a given detector at time t and position (x,y) is determined by the degree to which neighbouring regions are also active with respect to this detector and the activity of others.

(T_3) The associativity between detector arrays (i,j) is adaptive to their responses.

(T_4) A perceptual classification is made after this system of dynamically responding detectors reaches equilibrium.

(T_5) Classification of positional information into textured regions is accomplished by a weighted form of the minimum distance classifier, weighted by the total texture "entropy".

We have implemented all three proposed processes (convolution, impletion/cooperativity, and classification) to quantitatively observe the behaviour of the system with four critical texture pairs consisting of grey-scaled textures differing in

granularity, simple textures differing in orientation, and those differing in micropattern space characteristics: T,L, etc. We have chosen these latter two pairs since it has been (a) shown that they differ in discriminability, and (b) it has been proposed that they require "end-of-line" and intersection detectors to discriminate (Julesz, 1984)--the latter we can disprove. These are shown in Figure 1 together with the outputs of the classification procedure, using (13) to represent the relative "strength" of discrimination: Convergence usually occurred within 5-7 iterations.

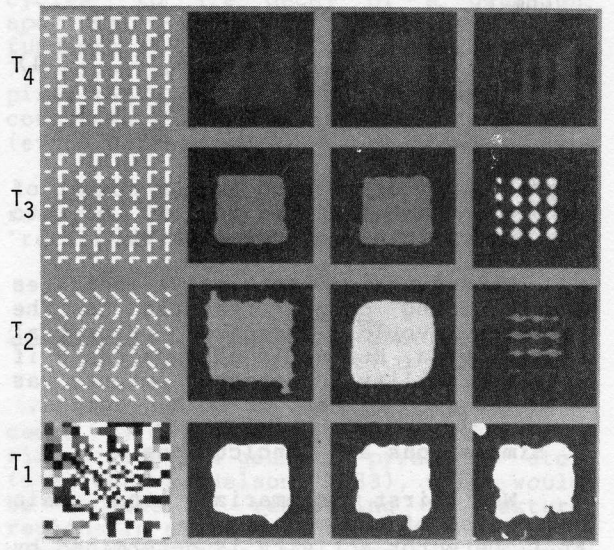

Figure 1.

Input textures (left column) and segmentation resulting from the outputs of 24 detectors with no associativities (second column), associativities as defined by (11) (column three) and inhibitory mass action (column four, equation 9): K=0.05 in (8). Contrast reflects segmentation strength according to equation (14).

To illustrate the effects of associativities on decreasing the dimensionality of the classification process, Figure 2 shows the eigenvalues for the solutions shown in columns two and three of Figure 1. Such reductions in "dimensionality" are clearly related to the iterative process converging to common strongly active detector profiles and inhibiting less active and isolated ones.

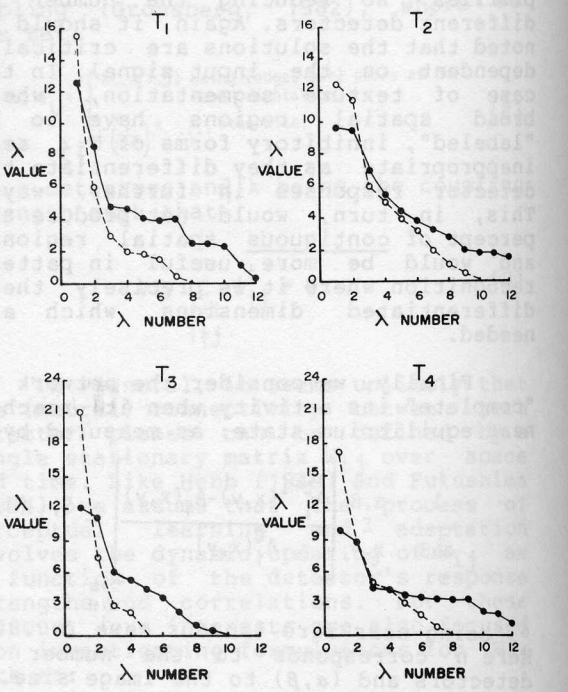

Figure 2.

Eigenvalues for the non-associative (solid lines: column 2, Figure 1) and asociative (dashed lines: column 3, Figure 1) segmentation processes. T_1 to T_4 correspond to the 4 textures shown up Column 1 of Figure 1.

What connects the texture processing equation (7) and the "texton" approach is that such cooperative networks decrease the dimensionality of the problem to the more strongly active--though "adaptively generated"--detectors or dimensions. That is, the profile of each detector in the process described by (8) is not stationary but, rather, is adapted by the energy it is designed to process and the activity of other units. Indeed, the actual profile at any time is recoverable by inverse filtering.

This model for texture segmentation is algebraically similar to a class of models for pattern recognition based on the associative (coupled) activities of large numbers of computational units whose activity profiles adapt to the signal and network states (see Kohonen, 1977; Fukushima, 1984). The main difference lies in how each computational component is interpreted, and the involvement of a classification scheme at the end, which actually produces the textured regions. In this sense the model is not formally dependent upon the initial edge and bar

detectors chosen, but rather on the ways in which their outputs are correlated over space and time according to (7).

In the present texture processing model the nature of the decomposition, and so dimensions, of a given texture segmentation task is dependent on the signal and the type of coupling operating between the computational units. If the visual system (or, indeed, the scientist) were to choose detectors which satisfied absolute orthogonality ($d_i * d_j \equiv 0$, * being convolution) then, from a mathematical perspective the ideal detector conditions would be present and the cooperative processes defined by (11) and (12) would not be required. However, the Impletion process would be involved, along with the classification algorithm. However, one assumption here is that the visual system is not that precise in creating detectors which, a priori, are so independent. Rather, the idea is that the visual system converges on the central detector profiles by adaptation processes like those described here--being signal dependent and network specific.

In conclusion, then, we have extended an earlier model for texture segmentation initially related to the "heuristics" of Julesz and Bergen (1983) for "preattentive" and "attentive" visual processes in spatial vision. The model has three components: decomposition (via cross-correlation), local and global processing, and classification. Though many questions still remain unanswered, our results suggest that these mechanisms, in a psychophysical sense, represent the types of processing involved in texture segmentation. The main result here is that the enumeration of detector profiles is but one part of the texture discrimination process and that the detector profiles "attended to" by the visual system are signal dependent and not fixed and invariant over all texture types, but also resultant from the underlying cooperative processes which generate "texton" classes to optimize the classification process by as few dimensions as possible. The proposed model does not solve the apparent rotation invariance processing characteristics of texture micropatterns, nor does it propose only one form of cooperative process. In this case we have found that the global process defined by (7) is efficient in reducing the dimensionality of the classification problem and have proposed that such coupling cannot be inhibitory if some form of "filling-in" is required.

References

Ahmed, N. & Rao, K. (1975). Orthogonal Transforms for Digital Signal Processing. Berlin: Springer.

Anderson, J.A., Silverstein, J.W., Ritz, S.A. & Jones, R.S. (1977). Distinctive features, categorical perception, and probability learning: some applications of a neural model. Psychological Review, 85, 413-451.

Burt, P.J. & Adelson, E.H. (1983). The Laplacian pyramid as a compact image code. IEEE Transactions on Communications, COM-31(4), 532-540.

Beck, J. (1983). A theory of textural segmentation. *In* Jack Beck (Ed.) Human and Machine Vision. New York: Academic Press, 1-38.

Caelli, T. (1982). On discriminating visual textures and images. Perception and Psychophysics, 31, 149-159.

Caelli, T. (1985). Three processing characteristics of visual texture segmentation. Spatial Vision, 1(1), 19-30.

Caelli, T. & Julesz, B. (1978). On perceptual analyzers underlying visual texture discrimination: Part 1. Biological Cybernetics, 28, 167-175.

Daugman, J. (1983). Six formal properties of two-dimensional anisotropic visual filters: Structural principles and frequency/orientation selectivity. IEEE Transactions on Systems, Man, and Cybernetics, SMC-13, 882-888.

Fukushima, K. (1984). A hierarchical neural network model for associative memory. Biological Cybernetics, 50, 105-113.

Grossberg, S. (1982). Studies of the Mind and Brain: Neural Principles of Learning, Perception, Development, Cognition and Motor Control. Boston: Reidel.

Harvey, L.O. & Gervais, M.J. (1978). Visual texture perception and Fourier analysis. Perception & Psychophysics, 24, 534-542.

Hebb, D.O. (1949). The Organization of Behavior, New York: John Wiley.

Hummel, R. & Zucker, S. (1983). On the foundations of relaxation labelling processes. IEEE: PAMI, 3, 267-287.

Julesz, B. (1981). Textons, the elements of texture perception and their interactions. Nature, 290, 91-97.

Julesz, B. (1984). Toward an axiomatic theory of preattentive vision. *In* Dynamic Aspects of Neocortical Function, G. Edelman, W. Gall, W.M. Cowan (Editors).

Julesz, B. & Bergen, J. (1983). Textons, the fundamental elements in preattentive vision and perception of textures. Bell Syst. Tech. Journal, 62, 1619-1645.

Kohonen, T. (1977). Associative memory--a system-theoretic approach. New York: Springer.

Richards, W. (1979). Quantifying sensory channels: Generalizing colorimetry to orientation and texture, touch and tones. Sensory Processes, 3, 207-229.