Texture Discrimination and Binding by a Modified Energy Model

Ko Sakai, Paul Sajda, and Leif H. Finkel
Department of Bioengineering and
Institute of Neurological Sciences
University of Pennsylvania
Philadelphia, PA. 19104-6392

Abstract

We describe a model of how textured surfaces are discriminated and represented by visual cortex. The model addresses two major processes: texture segmentation and texture binding. Textures are detected using a version of the energy model of Bergen, Adelson, and Landy [1] [2] which was modified to include ON and OFF center cells, and units selective for line endings. We describe a novel neural mechanism for binding a texture pattern together. Simulation results demonstrate the ability of the networks to segment and bind a well-known texture pattern.

Introduction

The visual system has a remarkable ability to discriminate subtle differences in texture patterns. Psychophysical studies have shown that texture discrimination occurs preattentively: namely, it operates in parallel over large regions of the visual field, occurs early in visual processing, and is unable to make distinctions based on multiple conjunctions of features [2] [9]. Recent approaches to understanding texture discrimination have followed two pioneering models. Julesz's texton model proposes that the visual system detects a relatively small number of primitive texture elements, called textons. Textons are features such as size, color, orientation, line endings, and junctions which, for the most part, are also the primitives to which visual cortical cells are selective. Julesz has shown that textures which differ in the density of one or more textons are distinguishable by human observers. In fact, textures which have identical second order statistics (i.e., identical Fourier transforms) and even identical third order statistics are still distinguishable if they differ in texton density.

A second recent approach to texture discrimination is the use of energy models [1]. The basic idea of these models is to sample, at several spatial frequencies, the amount of stimulus energy present (energy is loosely defined as the averaged squared output of a set of detecting elements). Such models are well-suited to network implementation and have been shown to work well in a variety of cases. The appeal of these models is that textures are discriminated based on the overall patterns of detector responses, rather than on differences in individual texture elements. In fact, individual elements are never even defined.

The output of either a texton or energy model is a segmentation, for example, the generation of a contour which separates a region from adjacent regions based on texture. However, texture perception involves more than segmentation, it also includes the generation of a *surface* and the *binding* of all texture elements on the surface together. The neural processes responsible for binding and surface representation are likely to be common to many visual processes, and could operate upon segmentations in motion, color, and depth as well as texture.

We have developed a model of how contours and surfaces are bound in the context of a model of depth-from-occlusion [3] [7]. We show here that a texture segmentation model can be alternatively used as input to this model, resulting in a disrimination of textured surfaces. In the next section, we describe this model, together with the extensions to Bergen and Landy's [2] energy model.

Construction of the Model

As shown in figure 1, the model is divided into networks concerned with *texture segmentation* and *texture binding*. The segmentation portion of the system consists of 38 interconnected, retinotopic maps, each containing 64x64 units. The network was simulated using the NEXUS neural simulation environment [8]. NEXUS is an interactive, window-based simulator which allows the anatomical and physiological properites of network units to be specified.

The input stimulus was presented as a 64x64 array. Following Bergen and Landy's model [2], we used networks selective for four orientations. However, our model differs from previous models in two respects. First, we incorporated both ON center and OFF center cells.[1] The second novel feature of our model is the inclusion of units sensitive to line-endings. We use a receptive field based on end-stopped cells in visual cortex. The addition of such units allows us to distinguish textures not discriminable based on orientation or size differences alone (see figure 3 for an example).
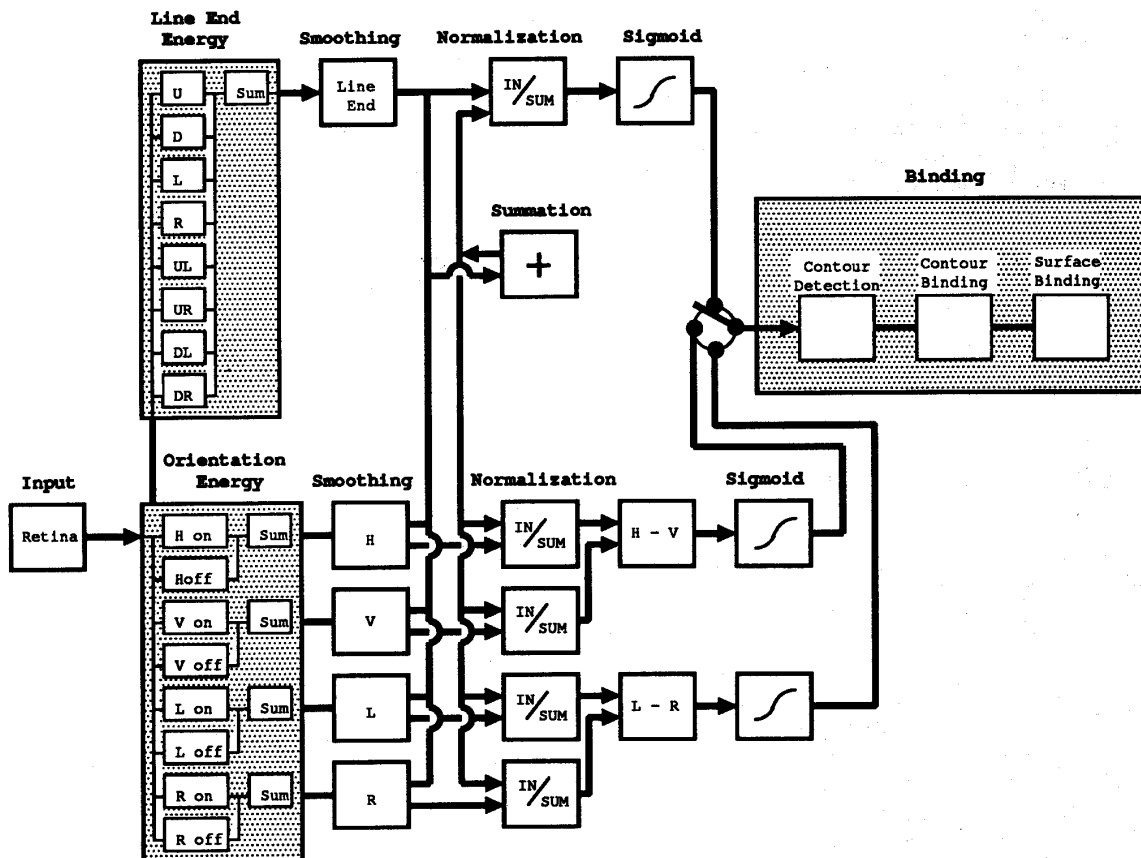

Figure 1: Computational flow of the simulation. Simulation consists of two major stages, texture segmentation and texture binding. Stimulus image serves as input to orientation and line-ending energy networks. Energies are locally averaged and then normalized by the total energy at each location. Orientation energies are then subtracted to form opponent pairs. After passing through a sigmoidal compression function, energies serve as input to the texture binding stage. Binding consists of three major parts, contour detection, contour binding and surface binding.

As indicated in figure 1, outputs are summed for all line ending units, and for ON and OFF pairs of orientation units. These outputs are independently "smoothed" by averaging responses over a region that is twice the size of the individual elements (11x11). Responses are then normalized by the total energy of the region in order to cancel out effects due to contrast of the local elements. We then form opponent pairs

[1]The utility of such an addition was anticipated by Julesz [5], and makes intuitive sense since the gaps between texture elements contain significant amounts of information. One can imagine two texture patterns with identical elements which differ only in the spacing between the elements. Since detector outputs are squared in an energy model, it is possible to use a single type of unit which responds positively to increased luminance and negatively to decreased luminance[2].

of horizontal-vertical and 45°-135° oriented units, and pass the results through a compressive sigmoidal nonlinearity which converts the graded analog responses to an approximately binary output.

This process of texture analysis was carried out at three different spatial frequencies. The receptive field masks for orientation and line ending units spanned 5x5 units, and we used Bergen and Landy's approach of shrinking (by 50%) and bluring the image to simulate the operation of detectors at different spatial scales. Only results from the highest spatial frequency units are shown below, as these units were most sensitive to the texture differences in the stimulus considered here.

A textured region is discriminated by the network when the responses to that region in one or more maps differ from responses to the surround. Bergen and Landy [2] identified the salient maps by hand, and then applied a directional derivative operator to generate a contour surrounding the texture region. We have used a more intrinsic approach by feeding back the outputs of the energy model to low-level orientation tuned cells. These cells, labelled "contour detection" in figure 1, pick up the edges of the texture region. Once this segmentation contour is generated, a *contour binding* network determines whether the contour is continuous and closed and, if so, links all units responding to the contour with a common "tag". We do not provide an explicit biophysical mechanism for such a tag, but recent results suggest that phase-locked cortical oscillations may play such a role[4].
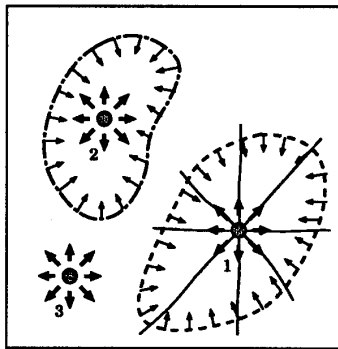

Figure 2: Schematic of the texture binding mechanism.

Figure 2 shows a schematic example of two texture contours passed to the texture binding system. Networks first determine the "inside" or *direction of figure* of each contour–this process identifies the direction of the surface along the contour. We have developed a novel neural mechanism by which units in the surface binding network bind all points belonging to the same textural surface. Shown are three surface binding units, each of which projects its dendrites in a stellate pattern (dendrites only shown explicitly for unit 1). Associated with each connection is a direction (shown with bold arrows). A particular connection is activated if it intersects a contour and if the direction of the connection is roughly opposite to the direction of figure at the site of the intersection (dot product of vectors is negative). For example, all connections for unit 1 are active since all connections have directions opposite (within +/- 90 degrees) to the local direction of figure. A connection is *not* activated if it either fails to intersect a contour, or if it proximally intersects a contour where its direction is the *same* as the direction of figure and then distally intersects the same contour where the connection direction and direction of figure are *opposite*. For example, the "north" connection for unit 3 would not be activated bu the dashed-dot contour. All active connections propagate the binding tag at the intersection to the unit. All units begin with the same initial tag and then each unit sets its activity to the tag propagated by the largest number of its connections. For example, unit 1 would have the tag associated with the dashed contour, unit 2 would have the dashed-dot tag. Since unit 3 has no active connections, it maintains its initial tag. All units compute their tags in parallel, resulting in units belonging to the same textural surface having the same tag. This spatial binding creates a surface and the following simulation demonstrates how such a texture surface is created.

Simulation Results

Figure 3 shows an example, adapted from Julesz [5], which contains two different textures. The central region contains "arrow" figures, and the surround contains triangles—both sets of elements are randomly oriented at 45° increments. The elements in the central square region differ from those in the surround only in the number of line endings present (note that the elements contain lines of the same lengths and orientations). It is actually somewhat difficult to recognize the shape of the central region without using focal attention. This may suggest that line endings are not as strong a segmentation cue as orientation or size.

A discretized version of this stimulus was presented to the network. As shown in figure 4, the early networks respond to the orientations and line endings in the stimulus. The texture difference is most strongly detected by the high spatial frequency units (outputs of lower frequency units are not displayed). As can be seen from the outputs of the sigmoidal compression maps, units selective for line endings clearly distinguish the two texture regions. The orientation selective maps (HV, LR) do not pick up the texture difference, but rather, respond to the random local structure of the line elements (in the lower spatial frequency maps (not shown), these patterns are not seen). Outputs of the "sigmoid" line ending map was fedback to orientation selective units which detect the edges of the central "square" texture region. The contour binding, direction of figure, and surface binding networks (not shown) then operate to bind the texture pattern into a surface.

The plot at the bottom of figure 4 displays the output of the surface binding network. For computational convenience, we have used the activity of units in this map to represent the binding tag associated with each unit (we do not believe that activity *per se* plays such a role *in vivo*). As can be seen, all units within the central texture region are bound with the same tag, and all units outside the region are bound with a separate, common tag. In addition, the contour surrounding the central region is bound with the same tag as the interior region, thus demonstrating that the contour is "owned" by the interior surface [3] [6].
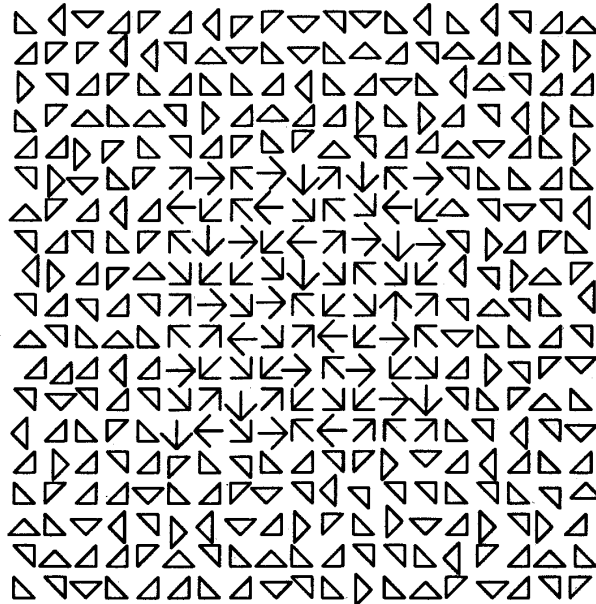
Figure 3: Textural Discrimination. Stimulus, adapted from Julesz [5] contains two texture regions: "arrows" in center and "triangles" in periphery. Elements are randomly oriented in 8 directions. Line orientations are same, on average, but number of line-endings differ in the two regions.
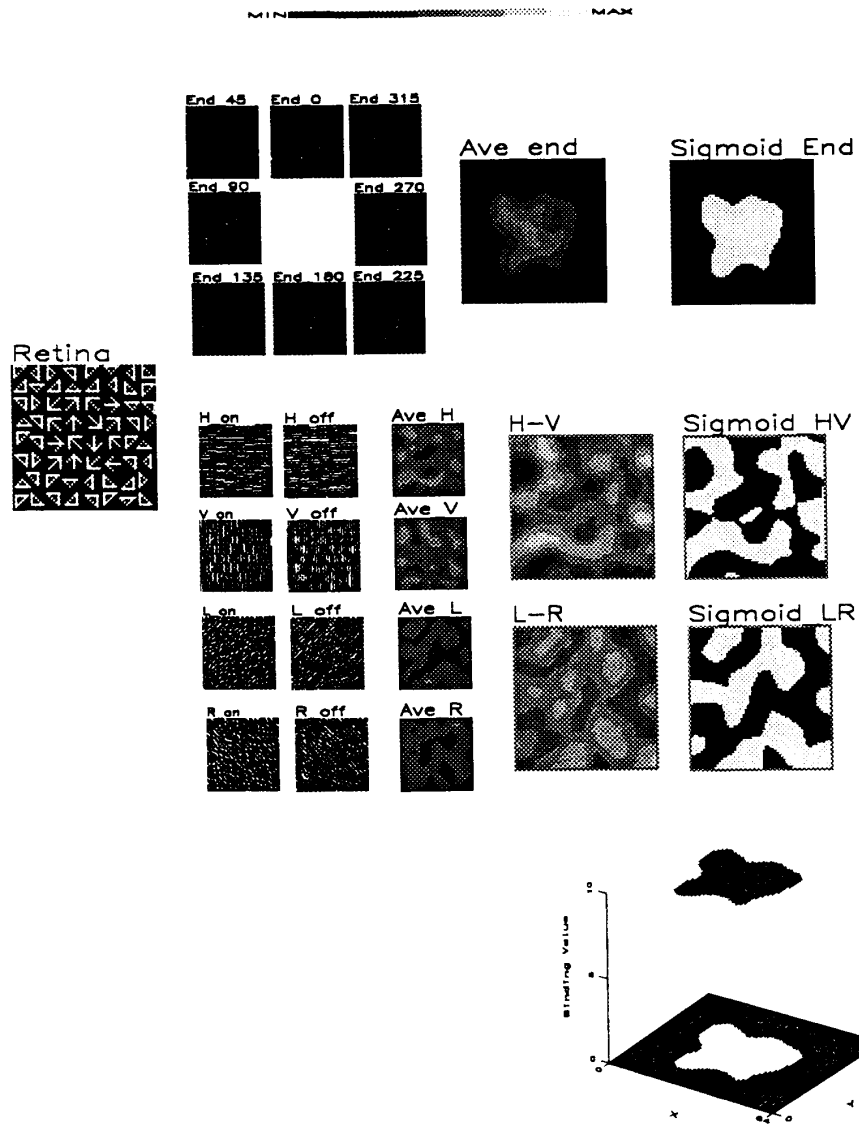
Figure 4: (Above) Results of network simulation. Discretized version of figure 3 was used as input. Responses are displayed for ON and OFF-center orientation selective networks (H on, H off, ...) and locally averaged and normalized orientation energy networks (Ave H, Ave V, ...). Responses of the eight oriented line-ending detection networks (End 0, End 45, ...) are also shown. "Ave end" network shows locally averaged and normalized energy of line-ending networks. "H-V" and "L-R" are opponent pairs of orientation energies. Networks at right labelled "sigmoid" show effect of sigmoidal compression. (Below) Three dimensional plot of units bound by texture binding system. Z-axis corresponds to the "tag" associated with the texture. In this simulation a center region is discriminated as having a different texture from the periphery, and units in these two regions are bound separately.

Conclusions

We have presented a model of how textured regions can be discriminated and textured surfaces created. The discrimination portion of the simulation uses a modified version of the energy model of Bergen, Adelson, and Landy[1] [2] which includes ON and OFF center cells and units selective for line endings. Surfaces are generated by a binding process in which all units inside the texture boundary contour are bound together (and to the contour).

It is clear that a wider range of textures could be discriminated by the inclusion of units selective for additional features. A reasonable collection might include the textons defined by Julesz [5], i.e., junctions or line crossing, luminance contrast or color, and units selective for aspect ratio (height/width). Separate units are not necessarily required for each of these features; for example, endstopped units may detect line length, curvature, and line terminations.

Our emphasis has not been on the initial segmentation of the visual scene into separately textured regions, but on the higher-order process of defining a textured surface. We have proposed that three processes are involved in surface representation: (1) determination of the inside of the boundary contour, (2) identification of all units responding to features within the texture region, and (3) binding of these units together with those responding to the contour. Additional attributes of the surface (depth, color, motion) would be similarly bound to it by other cortical modules.

Acknowledgements

References

[1] J. R. Bergen and E. H. Adelson. Visual texture segmentation and early vision. *Nature*, 333:363–364, 1988.

[2] J. R. Bergen and M. S. Landy. Computational modeling of visual texture segregation. In M. S. Landy and J. A. Movshon, editors, *Computational models of vision processing*, pages 253–271, MIT Press, Cambridge, 1991.

[3] L. H. Finkel and P. Sajda. Object discrimination based on depth-from-occlusion. *Neural Computation*, submitted, 1992.

[4] C. M. Gray and W. Singer. Neuronal oscillations in orientation columns of cat visual cortex. *Proceedings of the National Academy of Science USA*, 86:1698–1702, 1989.

[5] B. Julesz. Toward an axiomatic theory of preattentive vision. In G. M. Edelman, W. E. Gall, and W. M. Cowan, editors, *Dynamic aspects of neocortical function*, pages 585–612, John Wiley, New York, 1984.

[6] K. Nakayama and S. Shimojo. Toward a neural understanding of visual surface representation. *Cold Spring Harbor Symposia on Quantitative Biology*, LV:911–924, 1990.

[7] P. Sajda and L. Finkel. A neural network model of object segmentation and feature binding in visual cortex. *Int. Joint Conf. Neural Networks*, submitted, 1992.

[8] P. Sajda and L. Finkel. NEXUS: A simulation environment for large-scale neural systems. *SIMULATION, submitted*.

[9] A. Treisman and G. Gelade. A feature-integration theory of attention. *Cognitive Psychology*, 12:97–136, 1980.