Pre-attentive segmentation in the primary visual cortex

Zhaoping Li
Gatsby Computational Neuroscience Unit
University College London
17 Queen Square, London, WC1N, 3AR, U.K.
email: z.li@ucl.ac.uk
fax: 44 171 391 1173

Short title: Pre-attentive segmentation in V1.

Published in *Spatial Vision*, Volumn 13(1), p. 25-50, (2000).

Abstract

The activities of neurons in primary visual cortex have been shown to be significantly influenced by stimuli outside their classical receptive fields. We propose that these contextual influences serve pre-attentive visual segmentation by causing relatively higher neural responses to important or conspicuous image locations, making them more salient for perceptual pop-out. These locations include boundaries between regions, smooth contours, and pop-out targets against backgrounds. The mark of these locations is the breakdown of spatial homogeneity in the input, for instance, at the border between two texture regions of equal mean luminance. This breakdown causes changes in contextual influences, often resulting in higher responses at the border than at surrounding locations. This proposal is implemented in a biologically based model of V1 in which contextual influences are mediated by intra-cortical horizontal connections. The behavior of the model is demonstrated using examples of texture segmentation, figure-ground segregation, target-distractor asymmetry, and contour enhancement, and is compared with psychophysical and physiological data. The model predicts (1) how neural responses should be tuned to the orientation of nearby texture borders, (2) a set of qualitative constraints on the structure of the intracortical connections, and (3) stimulus dependent biases in estimating the locations of the region borders by pre-attentive vision.

1 Introduction

In early stages of visual processing, individual neurons respond directly only to stimuli in their classical receptive fields (CRFs)(Hubel and Wiesel, 1962). These CRFs sample the *local* contrast information in the input but are too small to cover visual objects at a *global* scale. Recent experiments show that the responses of primary cortical (V1) cells are significantly influenced by stimuli nearby and beyond their CRFs (Allman et al 1985, Knierim and Van Essen 1992, Gilbert, 1992, Kapadia et al 1995, Sillito et al 1995, Lamme, 1995, Zipser et al 1996, Levitt and Lund 1997). These contextual influences are in general suppressive and depend on the relative orientations of the stimuli within and beyond the CRF (Allman et al, 1985, Knierim and Van Essen 1992, Sillito et al 1995, Levitt and Lund 1997). In particular, the response to an optimal bar in the CRF is suppressed significantly by similarly oriented bars in the surround — iso-orientation suppression (Knierim and Van Essen 1992). The suppression is reduced when the orientations of the surround bars are random or different from the bar in the CRF (Knierim and Van Essen 1992, Sillito et al 1995). However, if the surround bars are aligned with the optimal bar inside the CRF to form a smooth contour, then suppression becomes facilitation (Kapadia et al 1995). The contextual influences are apparent within 10-20 ms after the cell's initial response (Knierim and Van Essen 1992, Kapadia et al 1995), suggesting that mechanisms within V1 itself are responsible (see discussion later on the different time scales observed by Zipser et al 1996). Horizontal intra-cortical connections linking cells with non-overlapping CRFs and similar orientation preferences have been observed and hypothesized to be the neural substrate underlying these contextual influences (Gilbert and Wiesel, 1983, Rockland and Lund 1983, Gilbert, 1992). There have also been theoretical studies of the mechanisms and phenomena of the contextual influences (e.g., Somers et al 1995, Stemmler et al 1995). However, insights into the computational roles of contextual influences have been limited to mainly contour or feature linking (Allman et al 1995, Gilbert, 1992, see more references in Li 1998a).

We propose that contextual influences serve the goal of pre-attentive visual segmentation by giving higher neural responses to potentially important locations in the input image, making these locations perceptually more salient. We call these relatively higher responses and the corresponding image locations, highlights. They can be caused by boundaries between texture or luminance regions, smooth contours, conspicuous targets, or outliers against backgrounds. This proposal will be demonstrated in a simple but biologically-based model of V1 with intracortical interactions between cells that are purely *local* (*ie* within the distance of a few CRFs). Note that although the horizontal intra-cortical connections are called long-range, they are still local with respect to the whole visual field since the axons reach only a few millimeters, or a few hypercolumns or CRF sizes, away from the pre-synaptic cells. Note also that finding the boundaries of regions or locating targets against backgrounds are two of the most essential components of segmentation, however they do not entail classifying or identifying the regions or targets. In other words, highlighting the important locations serves to process the "where" but not "what" of the underlying objects. We therefore propose that segmentation at its pre-attentive bare minimum is segmentation without classification (Li 1998b), i.e., segmentation without explicitly knowing the feature contents of the regions (see later discussion). This simplifies segmentation, making it plausible that it can be performed by low level, pre-attentive, processing in V1. This paper accordingly focuses on pre-attentive segmentation — additional processing is required to improve the resulting segmentation, e.g., by refining the coarse boundaries detected at the pre-attentive stage, classifying the contents of the regions, and segmenting in difficult cases when regions or targets do not pop out pre-attentively.

2 The principle and its implementation

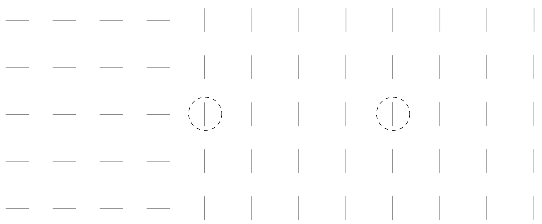
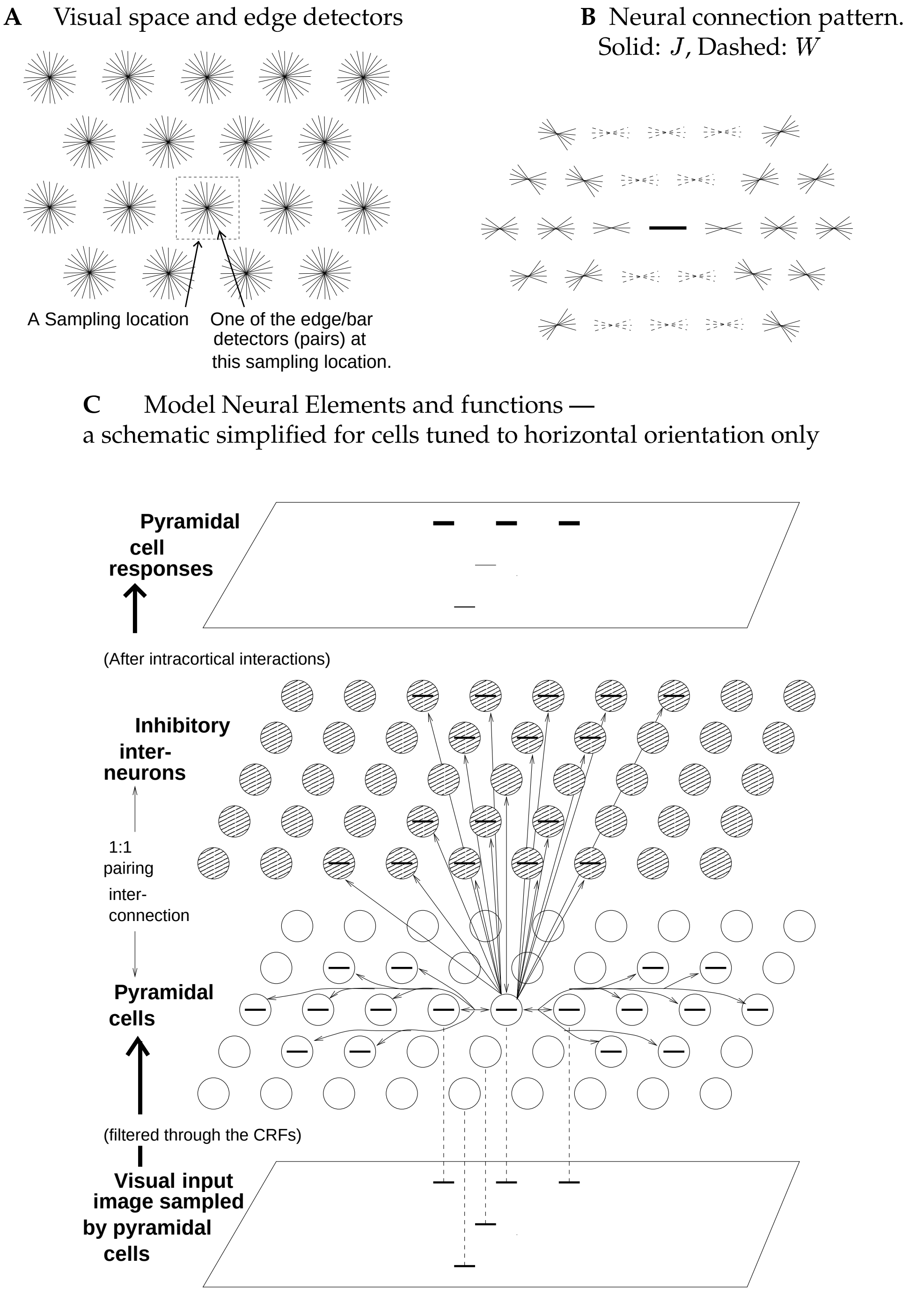


Figure 1: An input image, and two examples of CRFs marked by two dashed circles, which have the same stimulus within the CRFs but different contextual surrounds. A cell tuned to vertical orientation experiences less iso-orientation suppression in the left CRF than the right one.

The essential computational principle underlying our model is straightforward — detecting important image locations by detecting the breakdown of homogeneity or translation invariance in inputs. Consider a case in which the input is a large and homogeneous texture of equally and regularly spaced identical small bars. By symmetry, the responses of cells whose CRFs cover one part of the texture will be no different from responses of cells that cover another part, provided that the tuning properties of the cells are the same from one part to another, and that there is no spontaneous pattern formation (see later). Hence, no local input area will appear more salient than any other. However, if the texture patch is of finite size, the response of a cell depends on whether its CRF is near to or far away from the texture border, even if the contents of the CRF are exactly the same. The contextual surround lies wholly within the single texture region when the CRF is far from the border, but will include locations *outside* the texture region when the CRF is near the border (see Fig. (1)). The different surrounds make for different stimuli in the 'extra-classical' receptive field, i.e., different contextual influences, and consequently different cell responses. By the same argument, different cell responses are expected wherever one moves from one homogeneous region to another, i.e., wherever the homogeneity in input breaks down or the input characteristics change. A homogeneous region could be a blank region, a regular or even stochastic texture patch, or a region of characteristic input noise. A small target or smooth contour against a texture or noisy backgrounds also breaks the image homogeneity. It is these nonhomogeneous parts of the images that are usually of particular importance. Within the constraints of the existing experimental data, we construct in our model intra-cortical interactions (see below) such that the activities of neurons near region boundaries or isolated smooth contours will be relatively higher. This makes these locations relatively more salient, allowing them to pop out perceptually, thereby being pre-attentively segmented. Experiments in V1 indeed show that only 10-15 msec after the initial cell responses, activity levels are robustly higher near or at simple tex-



A Visual space and edge detectors

B Neural connection pattern. Solid: J, Dashed: W

C Model Neural Elements and functions — a schematic simplified for cells tuned to horizontal orientation only

A Sampling location
One of the edge/bar detectors (pairs) at this sampling location.
Pyramidal cell responses
(After intracortical interactions)
Inhibitory inter-neurons
1:1 pairing inter-connection
Pyramidal cells
(filtered through the CRFs)
Visual input image sampled by pyramidal cells

Figure 2:

ture boundaries or segments of a contour than inside homogeneous regions (Nothdurft, 1994, Gallant et al 1995, Kapadia et al 1995).

Our model focuses on segmentation in the absence of cues from color, motion, lu-

Figure 2: (Caption for figure in previous page) **A:** Visual inputs are sampled in a discrete grid by edge/bar detectors, modeling CRFs for V1 layer 2-3 cells. Each grid point has 12 neuron pairs (see **C**), one per bar segment. All 12 pairs or 24 cells at a grid point share the same CRF center, but are tuned to different orientations spanning 180^o, thus modeling a hypercolumn. A bar segment in one hypercolumn can interact with another in a different hypercolumn via monosynaptic excitation J or disynaptic inhibition W (see **B, C**). **B**: A schematic of the horizontal connection pattern from the center (thick solid) bar to neighboring bars within a finite distance (a few CRF sizes). J's contacts are shown by thin solid bars. W's are shown by thin dashed bars. Each CRF has the same connection pattern, suitably translated and rotated from this one. **C:** The neural elements and connections for cells tuned to horizontal orientations only (to avoid excessive clutter in the figure), and an illustration of the function of this system. Only connections to and from the central pyramidal cell are drawn. A horizontal bar, marking the preferred stimulus of the cell, is drawn on the central pyramidal cell and all other pyramidal or interneurons that are linked to it by horizontal connections. The central pyramidal sends axons (monosynaptic connections J) to other pyramidals that are displaced from it locally and roughly horizontally, and to the interneurons displaced locally and roughly vertically in the input image plane. The bottom plate depicts an example of input image containing 5 horizontal bars of equal contrast, each gives input to a pyramidal cell with the corresponding CRF (the correspondences are indicated by the dashed lines). These 5 pyramidal cells give higher response levels to the 3 bars aligned horizontally but lower responses to the 2 bars displaced vertical from them, as illustrated in the top plate. This is because the 3 horizontally aligned bars facilitate each other via the monosynaptic connections J, while the vertically displaced bars inhibit each other disynaptically via W.

minance, or stereo. Since it focuses on the role of contextual influences in segmentation, the model includes mainly layer 2-3 orientation selective cells and ignores the mechanism by which their CRFs are generated. Inputs to the model are images filtered by the edge- or bar-like local CRFs of V1 cells (we use 'edge' and 'bar' interchangeably). To avoid confusion, this paper uses the term 'edge' only for local luminance contrast, a boundary of a region is termed 'boundary' or 'border' which may or may not (especially for texture regions) correspond to any actual 'edges' in the image. Cells are connected by horizontal intra-cortical connections (Rockland and Lund 1983, Gilbert and Wiesel, 1983, Gilbert, 1992). These transform patterns of direct, CRF, inputs to the cells into patterns of contextually modulated output firing rates of the cells.

Fig. 2 shows the elements of the model and the way they interact. At each sampling location i there is a model V1 hypercolumn composed of cells whose CRFs are centered at i and that are tuned to 12 different orientations θ spanning 180^o (Fig. 2A). Based on experimental data (White, 1989, Douglas and Martin 1990), for each angle θ at location i, there is a pair of interconnected model neurons, an excitatory pyramidal cell and an inhibitory interneuron (Fig. 2C), so, altogether, each hypercolumn consists of 24 model neurons. Each model pyramidal cell or interneuron could model abstractly, say, 1000 pyramidal cells or 200 interneurons with similar CRF tuning (i.e., similar i and θ) in the real cortex, thus a 1:1 ratio between the numbers of pyramidal cells and interneurons in the model does not imply such a ratio in the cortex. For convenience, we refer to the cells tuned to θ at location i as simply the edge or bar segment $i\theta$.

Visual inputs are mainly received by the pyramidal cells, and their output activities (which are sent to higher visual areas) quantify the saliencies of their associated edge segments. The inhibitory cells are treated as interneurons. The input $I_{i\theta}$ to pyramidal cell $i\theta$ is obtained by filtering the input image through the CRF associated with $i\theta$. Hence, when the input image contains a bar of contrast $\hat{I}_{i\beta}$ at location i and oriented at angle β, pyramidal cells $(i\theta)$ are excited if β is equal or close to θ. The value $I_{i\theta}$ will be $\hat{I}_{i\beta}\phi(\theta-\beta)$ where $\phi(\theta-\beta) = e^{-|\theta-\beta|/(\pi/8)}$ describes the orientation tuning curve of the cell $(i\theta)$.

Fig. 2C shows an example in the case that the input image contains just horizontal bars. Only cells preferring orientations close to horizontal in locations receiving visual input are directly excited — cells preferring other orientations or other locations are not directly excited. In this example, the five horizontal bars have the same input strengths, and so the input $I_{i\theta}$ to the five corresponding pyramidal cells are of the same strengths as well. We omit cells whose preferred orientations are not horizontal but within the tuning width from horizontal for the simplicity of this argument.

In the absence of long-range intra-cortical interactions, the reciprocal connections between the pyramidal cells and their partner inhibitory interneurons would merely provide a form of gain control mechanism on input $I_{i\theta}$. The response from the pyramidal cell $i\theta$ would only be a function of its direct input $I_{i\theta}$. This would make the spatial pattern of pyramidal responses from V1 simply proportional to the spatial pattern of $I_{i\theta}$ up to a context-independent (*ie* local), non-linear, contrast gain control. However, in fact, the responses of the pyramidal cells are modified by the activities of nearby pyramidal cells via horizontal connections. The influence is excitatory via monosynaptic connections and inhibitory via disynaptic connections through interneurons. The interactions make a cell's response dependent on inputs outside its CRF, and the spatial pattern of response ceases being proportional to the input pattern $I_{i\theta}$.

Fig. 2B,C show the structure of the horizontal connections. Connection $J_{i\theta,j\theta'}$ from pyramidal cell $j\theta'$ to pyramidal cell $i\theta$ mediates monosynaptic excitation. Connection $J_{i\theta,j\theta'} > 0$ if these two segments are tuned to similar orientations $\theta \approx \theta'$ and the centers i and j of their CRFs are displaced from each other along their preferred orientation θ, θ'. Connection $W_{i\theta,j\theta'}$ from pyramidal cell $j\theta'$ to the inhibitory interneuron $i\theta$ mediates disynaptic inhibition from the pyramidal cell $j\theta'$ to the pyramidal cell $i\theta$. Connection $W_{i\theta,j\theta'} > 0$ if the preferred orientations of the two cells are similar $\theta \approx \theta'$, but the centers i and j of their CRFs are displaced from each other along a direction orthogonal to their preferred orientations. The reasons for the different designs of the connection patterns of J and W will be clear later.

In Fig. 2C, cells tuned to non-horizontal orientations are omitted to illustrate the intracortical connections without excessive clutter in the figure. Here, the monosynaptic connections J link neighboring horizontal bars displaced from each other roughly horizontally, and the disynaptic connections W link those bars displaced from each other more or less vertically in the visual input image plane. The full horizontal connection structure from a horizontal bar to bar segments including the non-horizontal ones is shown in Fig. 2B. Note that all bars in Fig. 2B are near horizontal and are within a distance of a few CRFs. The connection structure resembles a bow-tie, and is the same for every pyramidal cell within its ego-centric frame.

In the top plate of Fig. 2C, different bar widths are used to illustrate the different output activities in response to input bars of equal contrast. The three horizontally aligned bars in the input induce higher output responses because they facilitate each other's activities via the monosynaptic connections $J_{i\theta,j\theta'}$. The other two horizontal bars induce lower responses because they receive no monosynaptic excitation from others and receive disynaptic inhibition from the neighboring horizontal bars that are displaced vertically (and are thus not co-aligned with them). Note that the three horizontally aligned bars, especially the middle one, also receive disynaptic inhibitions from the two vertically displaced bars.

In the case that the input is a homogeneous texture of horizontal bars, each bar will receive monosynaptic excitation from its (roughly) left and right neighbors but disynaptic inhibition from its (roughly) top and bottom neighbors. Our intra-cortical connections are designed so that the sum of the disynaptic inhibition overwhelms the sum of the monosynaptic excitation. Hence the total contextual influence on any bar in an iso-orientation and homogeneous texture will be suppressive — iso-orientation suppression. Therefore, it is possible for the same neural circuit to exhibit iso-orientation suppression for uniform texture inputs and colinear facilitation (contour enhancement) for input contours that are not buried (i.e., obscured) in textures of other similarly oriented contours. This is exactly what has been observed in experiments (Knierim and van Essen 1992, Kapadia et al 1995).

To understand how a texture boundary induces higher responses, consider the two simple iso-orientation textures in Fig. (1). A bar at the texture boundary has roughly only half as many iso-oriented contextual bars as a bar in the middle of the texture. About half its contextual neighbors are oriented differently from itself. Since the horizontal connections only link cells with similar orientation preference, the contextual bars in the neighboring texture exert less or little suppression on the boundary bars. Therefore, a boundary bar induces a higher response because it receives less iso-orientation suppression than others. Similarly, one expects that a small target of one or a few bars will pop out of a homogeneous background of bars oriented very differently (e.g., orthogonally), simply because the small target experiences less iso-orientation suppression than the background bars. These intuitions are confirmed by later simulation results.

The neural interactions in the model can be summarized by the equations:

$$\begin{aligned} dx_{i\theta}/dt &= -\alpha_x x_{i\theta} - g_y(y_{i,\theta}) - \sum_{\Delta\theta\neq 0}\psi(\Delta\theta)g_y(y_{i,\theta+\Delta\theta}) + J_o g_x(x_{i\theta}) \\ &\quad + \sum_{j\neq i,\theta'} J_{i\theta,j\theta'}g_x(x_{j\theta'}) + I_{i\theta} + I_o \qquad (1) \\ dy_{i\theta}/dt &= -\alpha_y y_{i\theta} + g_x(x_{i\theta}) + \sum_{j\neq i,\theta'} W_{i\theta,j\theta'}g_x(x_{j\theta'}) + I_c \qquad (2) \end{aligned}$$

where $x_{i\theta}$ and $y_{i\theta}$ model the pyramidal and interneuron membrane potentials, respectively, $g_x(x)$ and $g_y(y)$ are sigmoid-like functions modeling cells' firing rates or responses given membrane potentials x and y, $-\alpha_x x_{i\theta}$ and $-\alpha_y y_{i\theta}$ model the decay to resting potentials, $\psi(\Delta\theta)$ is the spread of inhibition within a hypercolumn, $J_o g_x(x_{i\theta})$ is self excitation, and I_c and I_o are background inputs, including neural noise and inputs modeling the general and local normalization of activities (Heeger, 1992).

Depending on the visual stimuli, the system often settles into an oscillatory state (Gray and Singer, 1989, Eckhorn et al 1988), an intrinsic property of a population of recurrently connected excitatory and inhibitory cells. Temporal averages of the pyramidal outputs $g_x(x_{i\theta})$ over several oscillation cycles are used as the outputs, which coarsely model the pre-attentively computed saliencies of the stimulus bars. If the maxima over time of the responses of the cells were used instead as the model's outputs, the effects of differential saliencies shown in this paper would usually be stronger. That different regions occupy different oscillation phases could be exploited for segmentation (Li, 1998a), although we do not do so here. The complete set of model parameters to reproduce all the results in this paper are listed in Li (1998a), in which exactly the same model is used to account for contour enhancement (Polat and Sagi 1993, Field et al 1993, Kapadia et al 1995) rather than region segmentation.

The intra-cortical connections used in the model are consistent with experimental observations in that they tend to link cells preferring similar orientations, and that they synapse onto both the pyramidal cells and inhibitory interneurons (Rockland and Lund 1983, Gilbert and Wiesel, 1983, Hirsch and Gilbert 1991, Weliky et al 1995). To incorporate both co-linear excitation and iso-orientation suppression in the same neural circuit, we have assumed that the connection structure has an extra feature — the bow-tie — which is neither predicted nor contradicted by experiments. This extra feature correlates monosynaptic excitation or disynaptic inhibition with the degree of colinearity between two linked bars of similar orientations. We have found this structure to be necessary for the network to perform the required computation, and, as such, is a prediction of the model. We have also used dynamic systems theory to make sure that the system is well behaved. This imposes two particular constraints. First, colinear excitation has to be strong enough so that contours are enhanced, but not so strong as to excite bars which lie beyond the end of a contour but do not receive direct visual inputs. Second, the response to an iso-orientation homogeneous texture should also be homogenous, that is, the iso-orientation suppression should not lead to winner-take-all competition leading to the hallucination of illusory inhomogeneities within single regions (spontaneous pattern formation), so that no illusory borders occur within a single region. The same synaptic weights and all other model parameters, e.g., neural noise levels, cell threshold and saturation levels, are used for all simulated examples.

3 Performance of the model

The model was applied to a variety of input textures and configurations, shown in examples in figs (3 - 9). With a few exceptions (shown below), the input strengths $\hat{I}_{i\theta}$ are the same for all visible bars in each example so that any differences in the outputs $g_x(x_{i\theta})$ are solely due to the effects of the intra-cortical interactions. Using cell membrane time constants of the order of 10 msec, the contextual influences are significant after about 10 msec after the initial responses of the cells, agreeing with experimental observations (Knierim and van Essen 1992, Kapadia et al 1995, Gallant et al 1995).

The actual values $\hat{I}_{i\theta}$ used in all examples are chosen to mimic the corresponding experimental conditions. In this model the threshold and saturating input values are respectively $\hat{I}_{i\theta} = 1.0$ and $\hat{I}_{i\theta} = 4.0$ for an isolated input bar. Such a dynamic range, given an arbitrary scale for the threshold input, is comparable to physiological findings (Albrecht and Hamilton, 1982). Except for Fig. (9), all simulated examples in this paper use $\hat{I}_{i\theta} \geq 1.0$ for all visible bars plotted in the input images and $\hat{I}_{i\theta} = 0$ otherwise. Hence, we use $\hat{I}_{i\theta} = 1.05$ or 1.2 for low and near threshold input, and $\hat{I}_{i\theta} = 2.0$ and 3.5 for intermediate and high contrast input conditions used in experiments. Low input levels are used for all visible bars in Figs. (4B) and for the target bar in Fig. (3E, F, G, H) to demonstrate contour enhancement (Kapadia et al 1995, Kovacs and Julesz 1993). In-

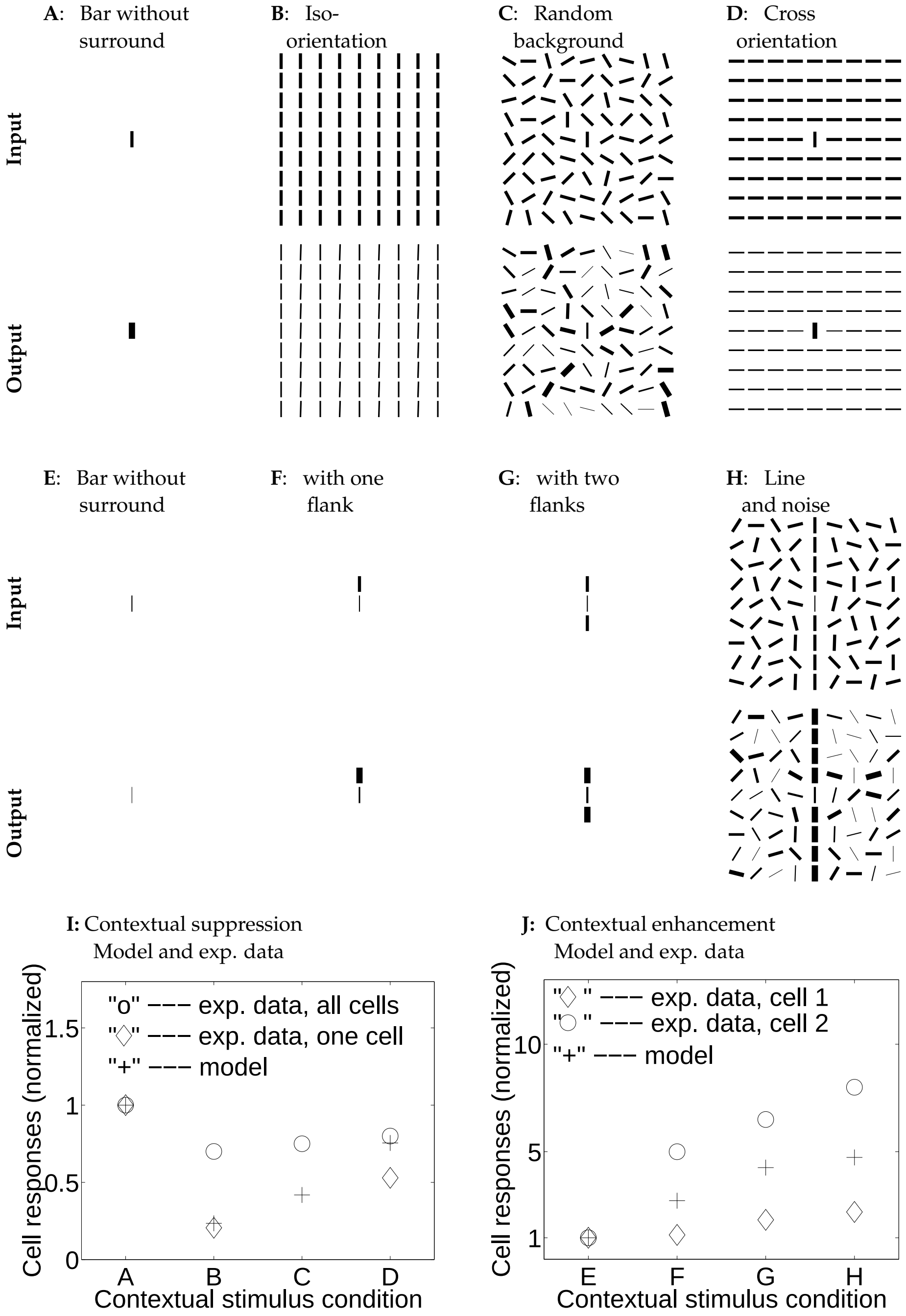


Figure 3:

termediate levels are used for all visible bars in texture segmentation and figure-ground pop-out examples (Figs. 4A, 5- 8). High input levels are used for all visible bars in Fig. (3A,B,C,D) and the contextual (background) bars in Fig. (3E,F,G,H) to model the high contrast conditions used in the physiological experiments that study contextual influence from textured and/or contour backgrounds (Knierim and van Essen 1992, Kapadia et al 1995). The output response strength $g_x(x_{i\theta})$ ranges in $[0, 1]$.

The plotted regions in all the figures are actually only small parts of larger images. In all cases, the widths of the bars in the figures are proportional to input or output strengths. For optimal visualization, the proportionality factor varies from figure to figure.

Fig. (3) shows that the model qualitatively replicates the results of physiological experiments on contextual influences from beyond the CRFs (Knierim and van Essen 1992, Kapadia et al, 1995). The suppression from surrounding textures in the model is strongest when the surround bars have the same orientation as the center bar, is weaker when the surround bars have random orientations, and is weakest when the surround bars have orthogonal orientations to the center bar. That the orthogonally oriented surround should lead to the weakest suppression is expected since the intra-cortical

Figure 3: (Caption for figure in previous page) Simulating the physiological experiments by Knierim and van Essen (1992) and Kapadia et al (1995) on contextual influences to compare the model behavior with the experimental data. The model input stimuli are composed of a vertical (target) bar at the center surrounded by various contextual stimuli. All the visible bars have high contrast input $\hat{I}_{i\theta} = 3.5$ except for the target bar in **E, F, G, H** where $\hat{I}_{i\theta} = 1.05$ is near threshold. The input and output strengths are proportional to the bar widths with the same proportionality factors (one for the input and another for the output) across different subplots for direct comparison. **A, B, C, D** simulate the experiments by Knierim and van Essen (1992) where a target bar is presented alone or is surrounded by contextual textures of bars oriented parallel, randomly, or orthogonal to it, respectively. The responses to the (center) target bars in **A, B, C, D** are, respectively, $0.98, 0.23, 0.41$ (averaged over different random surrounds), 0.74. **E, F, G, H** simulate the experiments by Kapadia et al (1995) where a low contrast (center) target bar is either presented alone or is aligned with some high contrast contextual bars to from a line with or without a background of randomly oriented high contrast bars. The responses to the target bars in **E, F, G, H** are, respectively, $0.07, 0.19, 0.30, 0.33$. Note that the response to the near threshold input target bar in **H** is much higher than that to the high contrast target bar in **B**. Contour enhancement also holds in **H** when all bars have high input values, simulating the psychophysics experiment by (Field et al 1993). **I:** Comparing model behavior in **A, B, C, D** (see the horizontal axis of the plot) with experimental data by Kinerim and van Essen (1992). The experimental data points "o" are adopted from the figure 11 in Knierim and van Essen (1992) which averaged over all recorded cells, whether they are orientation selective or not, while data points "$\Diamond$" are adopted from fig. 4B for a single cell in Knierim and van Essen (1992). **J:** Comparing the model behavior in **E, F, G, H** (see the horizontal axis of the plot) with experimental data from Kapadia et al. (1995). The data "o" and "$\Diamond$" are adopted from the two cell examples in the Figure 12B, C in Kapadia et al (1995). The model behavior depends quantitatively on the input contrasts. In both **I** and **J**, the plotted cell responses are normalized such that the responses to the isolated bar is 1.

interactions only link bars preferring similar orientations, and is also the neural basis for pop out of a target bar among homogeneous background bars of a different orientation. Note that the response to the target bar is lower in Fig. (3D) than in Fig. (3A). That is, pop-out is manifested by having the responses to the target being higher than the responses to the background. Pop-out is not dependent on the responses to the target being higher in the face of one background than the responses to the target in the face of a different background — a target pops-out against a blank background as well.

The relative degree of suppression in the model is quantitatively comparable to physiologically measurements of orientation selective cells that are sensitive to orientation contrast (Knierim and van Essen 1992; denoted by the '◊'s in Fig. (3I)). When averaged over all cell types (the 'o's in Fig. (3I)), the suppression observed experimentally is quantitatively weaker than that in our model. One possible explanation for this discrepancy could be that the results in our model apply only to those pyramidal cells that are orientation selective, whereas in the experiment many different cell types were probably recorded, including some that are interneurons and some that are not even orientation selective (Knierim and van Essen 1992). Using just the same neural circuit, Figs. (3E,F,G,H, J) compare contextual facilitation with the corresponding physiological data (Kapadia et al 1995). This facilitation is expected on the basis of our "bow-tie" like connection pattern. The quantitative degree of response enhancement is different for the different cells recorded in the experiment, and varies with the input contrast level in our model.

Given these results, we can then expect our model to perform appropriately on the sort of contour detection and pop-out tasks that are typically used in psychophysical studies (Fig. 4). Indeed, pop-out has recently been observed physiologically in V1 (Kastner et al 1997, Nothdurft et al 1998). Contour enhancement is built explicitly into the connection structure of the model. However, the contours can also be seen as where input homogeneity breaks down — here it is the statistical characteristics of the image noise in the background that are homogeneous.

Fig. 5 shows how texture boundaries become highlighted. Fig. 5A shows a sample input containing two regions. Fig. 5B shows the model output. Fig. 5C plots the responses (saliencies) $S(c)$ to the bars averaged in each column c in Fig. 5B, indicating that the most salient bars are indeed near the region boundary. Fig. 5D confirms that the boundary can be identified by thresholding the output activities using a threshold, $thresh = 0.5$, which is used to eliminates outputs that fire less strongly than the fraction $thresh$ of the highest output $\max_{i\theta}\{g_x(x_{i\theta})\}$ in the image. Note that V1 does not perform such thresholding, it is performed in this paper only for the purposes of display. The value of the threshold used in each example has been chosen for optimal visualization. To quantify the relative salience of the boundary, define the net salience at each grid point i as that of the most activated bar ($\max_\theta\{g_x(x_{i\theta})\}$), let S_{peak} be average salience across the most salient grid column parallel to and near the boundary, and $\bar{S}$ and σ_s be the mean and standard deviation in the saliencies of all locations. The relative salience of the boundary can be assessed by two quantities $r \equiv S_{peak}/\bar{S}$ and $z \equiv (S_{peak} - \bar{S})/\sigma_s$ (although these may be psychophysically incomplete as measures). r can be visualized from the thicknesses of the output bars in the figures, while z models the psychological z score. A salient boundary should give large values for (r, z). In Fig. (5), $(r, z) = (3.7, 4.0)$.

Note that the vertical bars near the boundary are more salient than the horizontal ones. This is because the vertical bars run parallel to the boundary, and are therefore specially enhanced through the contour enhancement effect of the contextual influences. This is related to the psychophysical observation that texture boundaries are

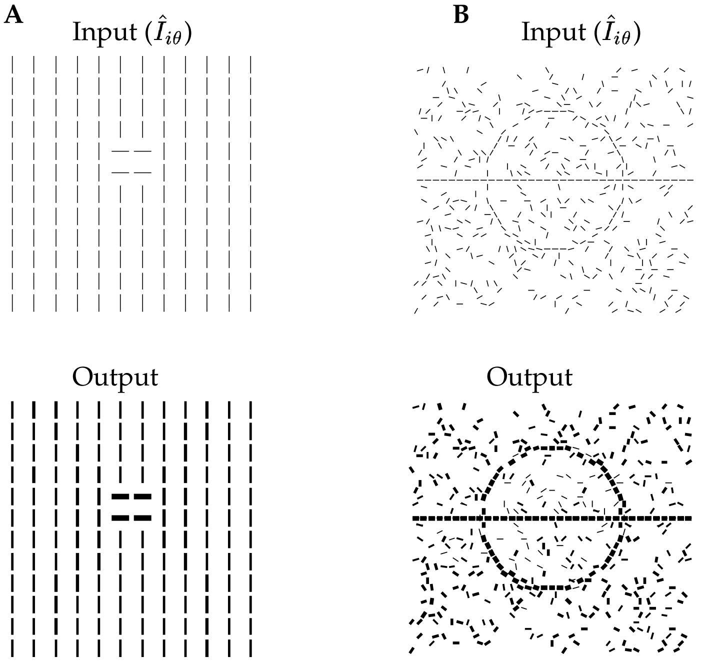
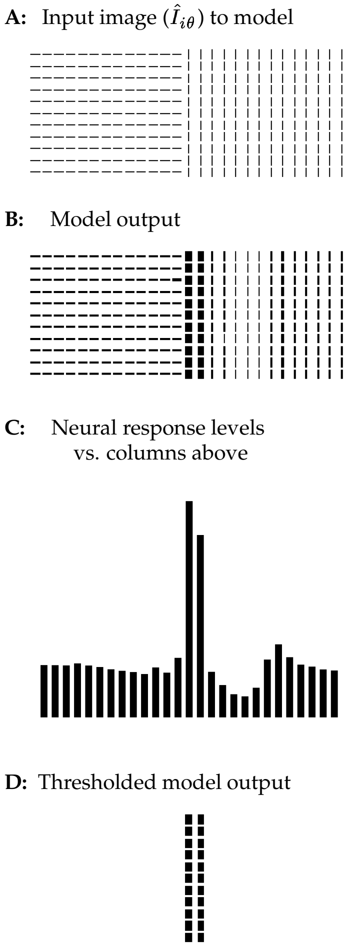


Figure 4: **A:** A small region pops out since all parts of it belong to the boundary. The response to figure is 2.42 times of the average response to the background. **B:** Exactly the same model circuit (and parameters) performs contour enhancement. The input strength is $\hat{I}_{i\theta} = 1.2$. The responses to the contour segments are 0.42 ± 0.03, and to the background elements 0.18 ± 0.08.

stronger when the texture elements on one side of them are parallel to the boundaries (Wolfson and Landy 1995), cf. Fig (6A) and Fig. (5). In fact, the model predicts that cells responding to bars near texture borders should be tuned to the orientation of the borders, and that the preferred border orientation should be the same as the preferred orientation of the bar within the CRF, as shown in Fig. (6L).

Fig (6) shows other examples demonstrating how the strength of the border highlight decreases with decreasing orientation contrast at the border, increasing orientation noise in the texture elements, or increasing spacing between the texture bars. When the orientation contrast is only 15^o, the boundary strength is very weak, with the boundary measures $(r, z) = (1.03, 0.78)$ (Fig.(6C)) — here the input is nearly homogeneous, making the boundary very difficult to detect pre-attentively. Stochasticity in the orientations of the bars also makes the border difficult to detect. In the example of Fig.(6F),



A: Input image ($\hat{I}_{i\theta}$) to model

B: Model output

C: Neural response levels vs. columns above

D: Thresholded model output

Figure 5: An example of the segmentation performance of the model. **A**: Input $\hat{I}_{i\theta}$ consists of two regions; each visible bar has the same input strength. **B**: Model output for **A**, showing non-uniform output strengths (temporal averages of $g_x(x_{i\theta})$) for the bars. **C**: Average output strengths (saliencies) in a column vs. lateral locations of the columns in **B**, with the heights of the bars proportional to the corresponding bar output strengthes. **D:** The thresholded output from **B** for illustration, $thresh = 0.5$.

an 90^o orientation contrast at the texture border is more severely smeared by orientation noise. A further result of the noise is that orientation contrasts are created at locations *within* texture regions, making them compete with the border for salience. These behaviors of the model can be understood in terms of the properties of the contextual influences. When the orientation contrast between two texture regions is small, a bar near the border receives near-orientation suppression from bars in the other texture region, weakening its response. With larger bar spacing or orientation noise, the overall iso-orientation suppression within a region is weaker, making less pronounced the difference in saliencies between bars near the border and bars within regions.

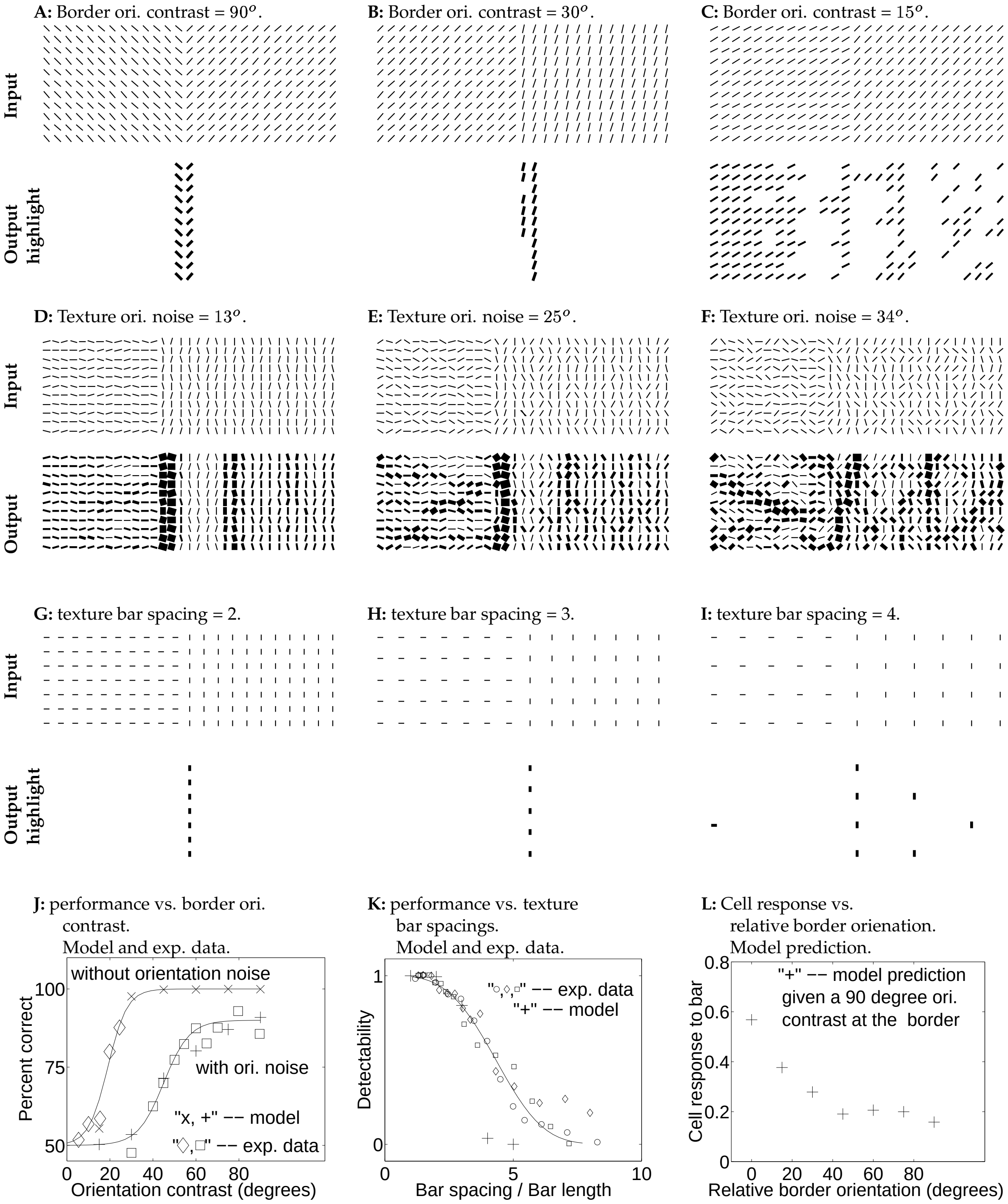


Figure 6:

In Fig.(6I), the texture elements are very sparse, and the boundary strengths are very weak $(r, z) = (1.02, 0.6)$. By comparison, the denser texture elements in Fig.(6H) give a boundary with $(r, z) = (1.1, 2.1)$. Even though the salience of the boundary is only 10%

Figure 6: (Caption for figure previous page) Model's segmentation performance, comparison with psychophysical data, and a prediction. **A, B, C, D, E, F, G, H, I:** Additional examples of model performance in different stimulus configurations. Each is an input image as in Fig. 5A followed immediately below by the corresponding thresholded (strongest) model outputs as in Fig. 5D, or unthresholded model output as in Fig. 5B. **A, B, C** show the effects of orientation contrasts at the border. **D, E, F** show the effects of orientation noise in regions. The neighboring bars within one texture region differ in orientation by a random amount whose averages are respectively 13^o, 25^o, and 34^o. **G, H, I** show the effects of texture bar spacing. **J**: segmentation performance versas orientation contrast at the border from the model, '+' and 'x', and psychophysical behavior '◊' and '□' from Nothdurft (1991). Data points '◊' and 'x' are for cases without orientation noise, '□' and '+' cases when the orientation noise amounts to 25^o on average between neighboring bars within one texture region (as in **E**). Given an orientation contrast, the data point '+' or 'x' for the model is an average of all possible relative orientations between the texture bars and the texture border. **K**: segmentation performance versas bar spacing from the model '+' and psychophysical behavior '◊', 'o', and '□' (each symbol for a particular bar length) from Nothdurft (1985). All data points are obtained from stimuli with a fixed orientation contrast 90^o at the border and without orientation noise. **L**: Cell responses near a texture border versas relative orientations between the border and the bars within the CRFs. The plot is based on a fixed orientation contrast 90^o at the border. The threshold used to obtain the output highlights in **A, B, C, G, H, I** are, respectively, $thresh = 0.77, 0.902, 0.8775, 0.92, 0.95, 0.935$.

higher than average, it has a high z score, i.e., the saliencies in the background are very homogeneous, making a 10% difference very noticable. By contrast, $(r, z) = (2.0, 1.4)$ for Fig.(6F) — although the average salience of the boundary column is 100% higher than the average, this difference is less significant with a z score $z = 1.4$ because the salience in the background is very inhomogeneous. To compare the performance of the network with psychophysical data (Nothdurft 1985, 1991), we model human behavior, in particular the measures of *Percentage of Correct Responses* or *Detectability* in locating or detecting texture borders, as functions of the z score of the boundary strength.[1] Fig (6J,K) compares the behavior of the model with psychophysical data on how segmentation performance depends on orientation contrast, orientation noise, and bar spacing.

Note also that the most salient location in an image may not be exactly on the bound-

[1] Let $P = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-t^2/2} dt$, be the probability of getting a measurement below z from a normal distribution. For psychophysical tasks in which subjects make binary judgements about the texture border, e.g., whether the border is oriented horizontally or vertically, the percentage of correct judgements is modeled by $Percent\ correct = (0.5 + P^n/2)100\%$, where n is a parameter used to account for various human, random, and task specific factors, such as the number of possible locations in an image for the texture border that a subject expects. For detectability tasks, such as detecting and identifyng the shapes of texture borders, the detectability in the model is modeled analogously by $Detectability = P^n$. This ranges in [0,1]. Parameter values $n = 18$ and $n = 60$ are used for the two conditions (with and without orientation noise) in Fig (6J); $n = 10$ is used for Fig (6K).

ary (Fig. 6B), or may be biased to one side of the boundary (Fig. (5B)). This should lead to a bias in the estimation of the border location, and can be experimentally tested. This is a hint that outputs from pre-attentive segmentation need to be processed further by the visual system.

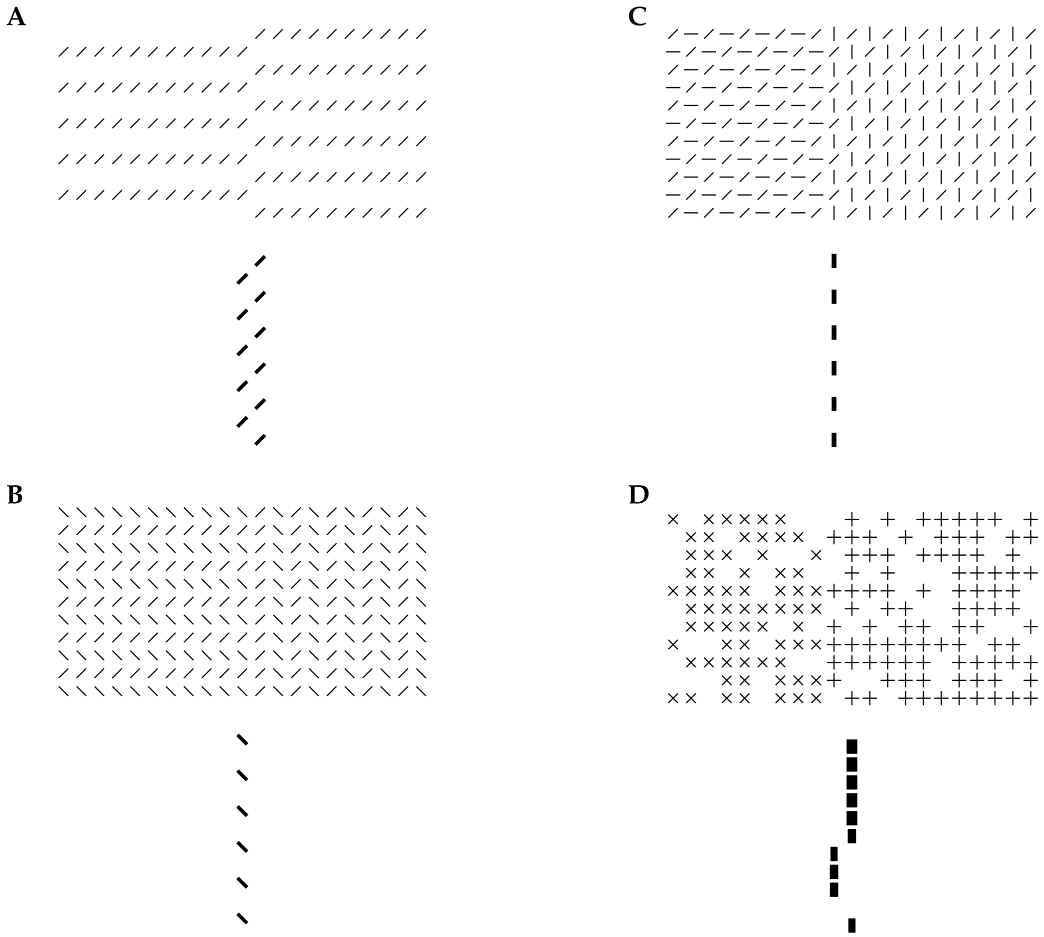


Figure 7: **A, B, C:** Model performance on regions with complex texture elements, and **D:** regions with stochastic texture elements. Each plot is the model input ($\hat{I}_{i\theta}$) followed immediately below by the output ($g_x(x_{i\theta})$) highlights. For **A, B, C, D** respectively, the thresholds to generate the output highlights are $thresh = 0.91, 0.9, 0.85, 0.56$.

The model also copes well with textures defined by complex or stochastic patterns (Fig. (7)). In both Figs. 7A and 7B, it segments the regions even though they have the same bar primitives and densities. In particular, the two regions in Fig. 7A have exactly the same features and would naturally be classified as the same texture. Segmentation in this case would be difficult if region boundaries were located by finding where texture classifications differ, as in traditional approaches to segmentation. Fig. 7C demonstrates that the model locates the correct border without being distracted by orientation contrast at locations within each texture region. The weakest boundary in Fig. (7) is that in Fig. (7B) with $(r, z) = (1.1, 1.5)$, and the other three cases have z scores close to or higher than $z = 3$.

The model works in these cases because it is designed to detect where input homogeneity or translation invariance breaks down. In the example of Fig. 7C, any particular vertical bar within the right region and far enough away from the border, has exactly the same contextual surround as any other similarly chosen vertical bar, i.e, they are all within the homogeneous or translation invariant part of the region. Therefore none of such vertical bars will induce a higher response than any other since they have the same direct input and the same contextual inputs. The same argument applies to the oblique bars or horizontal bars far away from the border in Fig. 7C as well as in Fig. 7A,B. However, the bars at or near the border do not have the same contextual surround (i.e., contextual inputs) as those of the other bars, i.e., the homogeneity is truely broken. Thus they will induce different responses. By design, the border responses will be higher. In other words, the model, with its translation invariant horizontal connection pattern, only detects where input homogeneity breaks down, and the pattern complexity within a region does not matter as long as the region is homogeneous. Orientation (or feature) contrasts often spatially coincide with the breakdowns of input homogeneity. However, they should not be taken as the *general* basis for segmentation. The stochasticity of the bars in Fig. 7D results in a non-uniform response pattern even within each region. In this case, just as in the examples in Fig. (6)D,E, the border induces the highest responses because it is where homogeneity breaks most strongly.

Our model also accounts for the asymmetry in pop-out strengths that is observed psychophysically (Treisman and Gormican, 1988), i.e., item A pops out among a background of items B more easily than vice versa. Fig. (8) demonstrates such an example, in which a cross among bars pops out much more readily than a bar among crosses. The horizontal bar in the target cross (among background vertical bars) experiences no iso-orientation suppression, while the vertical target bar among the crosses experiences about the same amount of iso-orientation suppression as the other vertical bars in the background crosses and therefore is comparatively weaker. This mechanism is effectively a neural basis for the psychophysical theory of Treisman and Gelade (1980) which suggests that targets pop out if they are distinguished from the background by possessing at least one feature (e.g., orientation) that the background lacks, but that they will not pop out if they are distinguished only by *lacking* a feature that is present in the background. Other typical examples on visual search asymmetry in the literature can also be qualitatively accounted for by the model (Li 1998b, 1999). Such asymmetry is expected in our framework — the nature of the breakdown in homogeneity in the input, i.e., the arrangement of direct and contextual inputs on and around the target features, is quite different depending on which image item is the figure and which is the background.

Fig. (9) shows the performance of the model on an input that is a photograph. The photo is sampled by the CRFs, and the outputs from even and odd simple cell CRF filters are combined to form phase insensitive edge/bar inputs to our model cells. Because of the sparse and single scale sampling of our current model implementation, the input to the model is rather degraded compared with the original photo. Nevertheless, the power of the model's intracortical processing is demonstrated by the fact that the

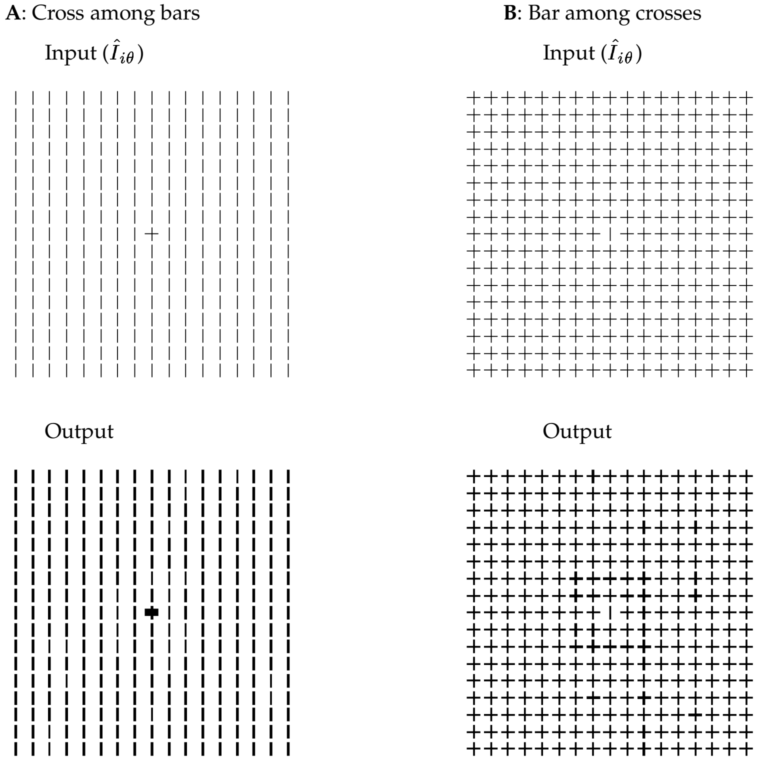


Figure 8: Asymmetry in pop-out strengths. **A**: The response to the horizontal bar in the cross is 3.4 times that to the average background. **B**: The area near the central vertical bar is the most salient part in the image, but the neural responses there are no more than 1.2 times that in the average background. The target bar itself is actually a bit less salient than the average background.

texture border is correctly highlighted (together with some other conspicuous image locations). It is important to stress that this paper isolates and focuses on the role of the intracortical computation – the model's performance, particularly on natural images, can only be improved using better methods of front-end processing, such as denser and multiscale sampling.

4 Summary and discussions

4.1 Summary of the results

We have presented a model which shows how contextual influences in V1 can play the computational role of mediating pre-attentive segmentation. The model's components and behavior are based on, and are consistent with, anatomical and physiological experimental evidence (Rockland and Lund, 1983, White, 1989, Douglas and Martin, 1990, Gilbert, 1992, Nothdurft, 1994, Gallant et al, 1995). This is the first model of V1 to cap-

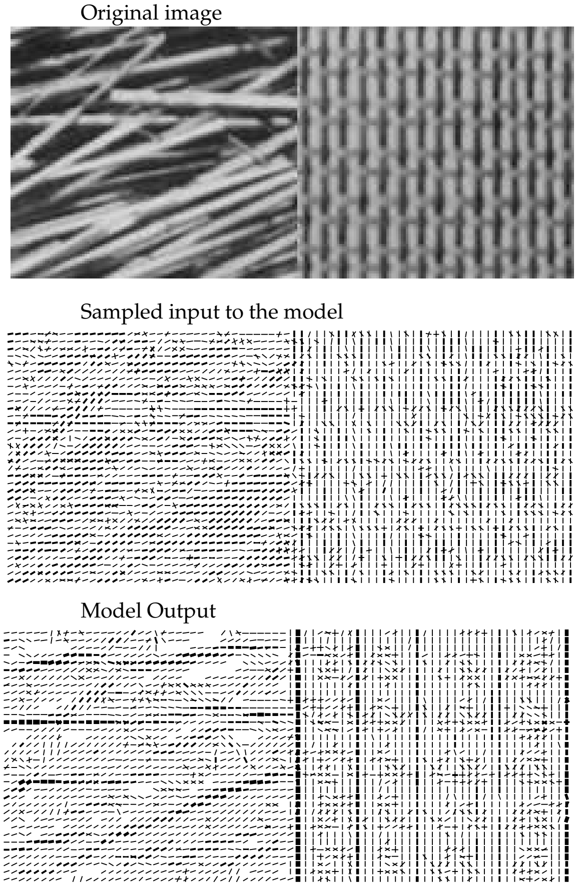


Figure 9: The model performance on a photo input. The photo is sampled by even and odd CRFs whose outputs are combined to form the phase insensitive edge/bar input signals to the model pyramidal cells. At each grid point, bars of almost all $K = 12$ orientations have nonzero input values $I_{i\theta}$. For display clarity, no more than 2 strongest input or output orientations are plotted at each grid point in model input and output. The model input is much degraded from the photo through this sampling because of the sparse and single scale sampling in the model. The locations of highest responses include the texture border and some conspicous locations within each texture region. The image space has wrap around boundary condition. Some highlights in the outputs away from the boundary are caused by the finite size (of the texture regions) effect.

ture the effect of higher neural activities near region boundaries, the pop-out of small figures against backgrounds, and asymmetries in pop-out strengths between choices of figure and ground. The mechanism underlying the model is the *local* intra-cortical interactions that modify individual neural activities in a manner that is sensitive to the contextual visual stimuli, detecting region boundaries by detecting the breakdown of homogeneity in the inputs. The model is also the first to use the same neural circuit for both the region boundary effect and contour enhancement — individual contours in a noisy or non-noisy background can also be seen as examples of the breakdown of homogeneity in inputs. Our model suggests that V1, as a saliency network, has substantially more computational power than is traditionally supposed.

4.2 Relation to other studies

It has recently been argued that much texture analysis and segmentation are performed at low levels of visual processing (Bergen, 1991). The observed ease and speed of solving many texture segmentation and target detection tasks have long led theorists to propose that the computations responsible for this performance must be pre-attentive (Treisman and Gelade 1980, Julesz 1981, Treisman and Gormician 1988). Correspondingly, many relevant models using low level and autonomous network mechanisms have been proposed.

To characterize and thus discriminate texture regions, some models use responses of image filters resembling the CRFs of the V1 cells (Bergen and Adelson 1988); others use less biologically motivated feature measures such as pixel correlations and histograms (Haralick and Shapiro, 1992).

Another class of models goes beyond the direct responses from the image filters and includes interactions between the filter or image units. The interactions in these models are often motivated by their cortical counterparts. They are designed to characterize texture features better and often help to capture context dependences and other statistical characteristics of texture features. Some of these models are based on Markov Random Field techniques (Haralick and Shapiro, 1992, Geman and Geman, 1984); others are closer to neurobiology. For instance, Caelli (1988) suggests a dynamic model in which the image filter units interact locally and adaptively with each other such that the ultimate responses converge to an "attractor" state characterizing the texture features. Interestingly, this dynamic interaction can make the feature outputs of a figure region dependent on the features of background regions, and was used (Caelli 1993) to account for examples of asymmetries between figure and ground observed by Gurnsey and Browse (1989). Malik and Perona's model (1990) contains neural based multiscale filter (feature) units which inhibit each other in a form of winner-take-all competition which gives a single final output feature at each image location.

Many of these models capture well much of the phenomenology of psychophysical performance. They share the assumption that texture segmentation first requires characterizing the texture features at each image location. These feature measurements are then compared at neighboring locations (either by explicit interactions in the model, or, implicitly, by some unmodeled subsequent processing stage) to locate boundaries between texture regions.

By contrast, a recent model by Nothdurft (1997) directly locates the boundary between neighboring texture regions (of oriented bars) via local nonlinear inhibitory interactions between orientation units tuned to similar orientations. Thus, borders of textures of oriented bars are located as locations of comparatively stronger responses or higher saliencies, just as in our model. Our model goes beyond Nothdurft's model by adopting a framework to include other conspicuous image locations where homogeneity in inputs breaks down and by explicitly including both intracortical excitation and inhibition. Consequently, the neural circuit in our model can, in addition, detect borders between complex textures, and account for other seemingly less related phenomena such as pop-out, asymmetries between figure and ground, and contour enhancement.

There are also various related models which focus on the computation of contour enhancement. Li (1998a) presents a detailed discussion of them and their relation to the present model. In particular, a model by Grossberg and coworkers (Grossberg and Mingolla 1985, Grossberg et al 1997) proposes a "boundary contour system" as a model of intra-cortical and inter-areal neural interactions within and between V1 and V2. The model aims to capture illusory contours which link bar segments and line endings. Our model is the only one to models both contour enhancement and region boundary highlights in the same neural circuit. Of course, its instantiation in V1 limits its power in some respects – it does not perform computations such as the classification and smoothing of region features and the sharpening of boundaries that are performed by certain other models (e.g., Malik and Perona 1990).

Since it locates conspicuous image locations without using filters that are specifically tuned to complex region features (such as the '+'s and 'x's in Fig. (7D)), our model reaches beyond early visual processing using such things as center-surround filters (Marr, 1982). While the early stage filters code image primitives (Marr, 1982), our mechanism should help in the representations of object surfaces. Since contextual influences are collected over whole neighborhoods, the model naturally accounts for the way that regions are defined by the statistics of their contents. This agrees with Julesz's conjecture of segmentation by image statistics (Julesz, 1962) without imposing any restriction that only the first and second order image statistics are important. Julesz's concept of textons (Julesz, 1981) could be viewed in this framework as any feature to which the particular intra-cortical interactions are sensitive and discriminatory. Given the way that it uses orientation dependent interactions between neurons, our model agrees with previous ideas (Northdurft, 1994) that (texture) segmentation is primarily driven by orientation contrast. However the emergent network behavior is collective and accommodates characteristics of general regions beyond elementary orientations, as in Fig. 7. The psychophysical phenomena of filling-in (when one fails to notice a small blank region within a textured region) could be viewed in our framework as the instances when the network fails to sufficiently highlight the non-homogeneity in inputs near the filled-in area.

Our pre-attentive segmentation is quite primitive. It merely segments surface regions, whether or not these regions belong to different visual objects. It does not characterize or classify the region properties or categories. In this sense, this pre-attention segmentation process is termed segmentation without classification (Li, 1998b). Hence, for example, our model does not say whether a region is made of a transparent surface on top of another surface, nor does the model facilitate the pop-out process based on categorical informations such as the target being the only "steep" item in an image (Wolfe and Friedman-Hill 1992).

Our model suggests that there might be experimental evidence that pre-attentive segmentation precedes (and is dissociated from) visual classification or discrimination. Recent experimental evidence from V1 (Lamme et al 1997, Zipser, private communication 1998) shows that the modulation of neural activities starts at the texture boundary and only later includes the figure surface. While the response modulations at the figure boundary take about 10-20 ms to develop after the initial cell responses, they take about 50 ms within the figure surface away from the boundary (Zipser et al 1996, Zipser, private communication, 1998). Further, some psychophysical evidence (Scialfa and Joffe 1995) suggests that information regarding (figure) target presence is available before information about the feature values of the targets. Also, V2 lesions in monkeys are shown to disrupt region content discrimination but not region border detection (Merigan et al 1993). These results are consistent with our suggestion. Furthermore, neural modulation in V1, especially those in figure surfaces (Zipser 1998, private communication), is strongly reduced or abolished by anaesthesia or lesions in higher visual areas (Lamme et al 1997), while experiments by Gallant et al (1995) show that activity modulation at texture boundaries is present even under anaesthesia.

Taken together, this experimental evidence suggests the following computational framework. Pre-attentive segmentation in V1 precedes region classification; region classification following pre-attentive segmentation commences in higher visual areas; the classification is then fed back to V1 in the form of top-down influences which allows the segmentation to be refined (for instance, by removing the bias in the estimation of the border location in the example of Fig. 6B); this latter segmentation refinement process might be attentive and can be viewed as segmentation by classification. Finally, the bottom-up and top-down loop can be iterated to improve both classification and segmentation. Top-down and bottom-up streams of processing have been studied by many others (e.g., Grenander 1976, Carpenter and Grossberg 1987, Ullman 1994, Dayan et al, 1995). Our model studies the first step in the bottom up stream, which initializes the iterative loop. The neural circuit in our model can easily accommodate top-down feedback signals which, in addition to the V1 mechanisms, selectively enhance or suppress neural activities in V1 (see examples in Li 1998a). However, we have not yet modeled how higher visual centers process the bottom up signals to generate the feedback.

4.3 Model predictions

The experimentally testable predictions of the model are: (1) cell responses should be tuned to the orientation of nearby texture borders (Fig. (6L)), and the preferred border orientation should be the same as that of the bar within the CRF; (2): the horizontal connection should have a qualitative "bow-tie" structure as in Fig. 2B, with a dominant monosynaptic excitation between cells tuned to similarly oriented and co-aligned bars and a dominant disynaptic inhibition between similarly oriented but not co-aligned bars; (3): there should be stimulus dependent biases in the border locations that could be estimated on the basis of the neural responses (e.g., Fig. 6B) or by pre-attentive vision.

Since the model is quite simplistic in the design of the connections, we expect that there will be significant differences between the model and physiological connections. For instance, two linked bars interact in the model either via monosynaptic excitation or disynaptic inhibition, but not both. In the cortex, two linked cells often interact via both excitation and inhibition, making the overall strength of excitation or inhibition dependent on the input contrast (e.g., Hirsch and Gilbert, 1991; see Li 1998a for analysis). Hence, the excitation (or inhibition) in our model could be interpreted as the abstraction of the predominance of excitation (or inhibition) between two linked bars. Currently, different sources of experimental data on the connection structure are not mutually consistent regarding the spatial and orientation dependence of excitation and inhibition (Fitzpatrick 1996, Cavanaugh et al 1997, Kapadia, private communication 1998, Hirsch and Gilbert 1991, Polat et al 1998). This is partly due to different experimental conditions, such as different input contrast levels and different stimulus elements (e.g., bars or gratings). The performance of the model is also quantitatively dependent on input strength. One should bear this fact in mind when viewing the comparisons between the model and experimental data in Figs. (3, 6).

4.4 Limitations and extensions of the model

Our model is still very primitive compared with the true complexity of V1. A particular lacuna is multiscale sampling. This is important, not only because images contain multiscale features, but also because arranging for the model to treat flat surfaces slanted in depth as “homogeneous” or “translation invariant” requires some explicit mechanisms for interaction across scales. Merely replicating and scaling the current model to multiple scales is not sufficient for this purpose. In addition to orientation and spatial location, neurons in V1 are tuned for motion direction/speed, disparity, ocularity, scale, and color (Hubel and Wiesel 1962, Livingstone and Hubel 1984), and our model should be extended accordingly. The intra-cortical connections in the extended model will link edge segments with compatible selectivities to scale, color, ocular dominance, disparity, and motion directions as well as orientations, as suggested by experimental data (e.g., Gilbert 1992, Ts’o and Gilbert 1988, Li and Li 1994). The extended model should be able to highlight locations where input homogeneity in depth, motion, color, or scale is broken.

Other desirable extensions and refinements of the model include the sort of dense and over-complete input sampling strategy that seems to be adopted by V1, more precisely determined CRF features, physiologically and anatomically more accurate intra-cortical circuits within and between hypercolumns, and other details such as on and off cells, cells of different CRF phases, non-orientation selective cells, end stopped cells, and more cell layers. These details should help to achieve a better quantitative match between the model and human vision.

Any given model, with its specific neural interactions, will be more sensitive to some region differences than others. Therefore, the model sometimes finds it easier or more difficult than humans to segment some regions. Physiological and psychophysical measurements of the boundary effect for different types of textures can help to constrain the horizontal connection patterns in an improved model. Experiments also suggest that the connections may be learnable or plastic (Karni and Sagi, 1991, Sireteanu and Rieth 1991, Polat and Sagi 1994).

We currently model salience at each location quite crudely, using just the activity of the single most salient bar. It is essentially an experimental question as to how the salience should best be defined, and the model can be modified accordingly. This will be particularly critical once the model includes multiple scales, non-orientation selective cells, and other visual input dimensions. The activities of cells in different channels need somehow to be combined to determine the salience at each location of the visual field.

In summary, this paper proposes that the contextual influences via intra-cortical interactions in V1 serve the purpose of pre-attentive segmentation. It introduces a simple, biologically plausible model which demonstrates this proposal. Although the model is as yet very primitive compared to the real cortex, our results show the feasibility of the underlying ideas, that simple pre-attentive mechanisms in V1 can serve difficult segmentation tasks, that breakdown of input homogeneity can be used to segment regions, that region segmentation and contour detection can be addressed by the same mechanism, and that low-level processing in V1 together with *local* contextual interactions can contribute significantly to visual computations at *global* scales.

Acknowledgement: I thank Peter Dayan for many helpful discussions and conversations over the duration of this research, he, John Hertz, Geoffrey Hinton, and John Hopfield for their careful readings and helpful comments on various versions of the paper, Sabina Wolfson for her careful readings and very helpful suggestions to significantly improve the presentation of this paper, a reviewer for his/her comments, many other colleagues for their questions and comments during and after my seminars, and the Center for Biological and Computational Learning at MIT for hosting my visit. This work is supported in part by the Hong Kong Research Grant Council and the Gatsby Foundation.

References

[1] Albrecht D. G. and Mamilton D. B. (1982). Striate cortex of monkey and cat: Contrast response function. *J. Neurophysiol.* Vo. 48, 217-237.

[2] Allman, J. Miezin, F. and McGuinness E. (1985) Stimulus specific responses from beyond the classical receptive field: neurophysiological mechanisms for local-global comparisons in visual neurons. *Ann. Rev. Neurosci.* **8**, 407-30.

[3] Bergen J.R. (1991) Theories of visual texture perception. In *Vision and visual dysfunction* D. Regan (Ed.) Vol. 10B (Macmillan),pp. 114-134.

[4] Bergen J. R. and Adelson, E. H. (1988) Early vision and texture perception. *Nature* **333**, 363-364.

[5] Caelli, T. M. (1988) An adaptive Computational Model for Texture Segmentation. *IEEE trans. Systems, Man, and Cybern.* Vol. 18, p9-17.

[6] Caelli, T. M. (1993) Texture classification and segmentation algorithms in man and machines. *Spatial Vision* Vol. 7 p 277-292.

[7] Carpenter, G & Grossberg, S (1987). A massively parallel architecture for a self-organizing neural pattern recognition machine. *Computer Vision, Graphics and Image Processing,* **37**, 54-115.

[8] Cavanaugh, J.R., Bair, W., Movshon, J. A. (1997) Orientation-selective setting of contrast gain by the surrounds of macaque striate cortex neurons *Soc. Neuroscience Abstract* 227.2.

[9] Dayan, P, Hinton, GE, Neal, RM & Zemel, RS (1995). The Helmholtz machine. *Neural Computation,* **7**, 889-904.

[10] Douglas R. J. and Martin K. A. (1990) Neocortex, in *Synaptic Organization of the Brain* G. M. Shepherd. (ed). Oxford University Press, 3rd Edition, pp389-438

[11] Eckhorn, R. Bauer R., Jordan W., Brosch M., Kruse W., Munk M., and Reitboeck H. J. (1988) Coherent oscillations: a mechanism of feature linking in the visual cortex? Multiple electrode and correlation analysis in the cat. *Biol. Cybern.* **60**, 121-130.

[12] Field D.J., Hayes A., and Hess R.F. (1993). Contour integration by the human visual system: evidence for a local 'associat ion field' *Vision Res.* 33(2): 173-93.

[13] Fitzpatrick D. (1996) The functional organization of local circuits in visual cortex: Insights from the study of tree shrew striate cortex. *Cerebral Cortex*, V6, N3, 329-341.

[14] Gallant, J. L. , van Essen, D. C. , and Nothdurft H. C. (1995) Two-dimensional and three-dimensional texture processing in visual cortex of the macaque monkey. In *Early vision and beyond* T. Papathomas, Chubb C, Gorea A., and Kowler E. (eds). MIT press, pp 89-98.

[15] Geman S. and Geman D. (1984) Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE trans PAMI* **6** 721-741.

[16] Gilbert, C. D. (1992) Horizontal integration and cortical dynamics *Neuron*. **9**(1), 1-13.

[17] Gilbert C.D. and Wiesel T.N. (1983) Clustered intrinsic connections in cat visual cortex. *J Neurosci.* **3**(5), 1116-33.

[18] Gilbert C. D. and Wiesel T. N. (1990) The influence of contextual stimuli on the orientation selectivity of cells in primary visual cortex of the cat. *Vision Res* 30(11): 1689-701.

[19] Gray C. M. and Singer W. (1989) Stimulus-specific neuronal oscillations in orientation columns of cat visual cortex *Proc. Natl. Acad. Sci. USA* **86**, 1698-1702.

[20] Grenander, U (1976-1981). *Lectures in Pattern Theory I, II and III: Pattern Analysis, Pattern Synthesis and Regular Structures.* Berlin: Springer-Verlag., Berlin.

[21] Grossberg S. and Mingolla E. (1985) Neural dynamics of perceptual grouping: textures, boundaries, and emergent segmentations *Percept Psychophys.* **38** (2), 141-71.

[22] Grossberg S. Mingolla E., Ross W. (1997) Visual brain and visual perception: how does the cortex do perceptual grouping? *TINS* vo. 20. p106-111.

[23] Gurnsey, R. and Browse R. (1989) Asymmetries in visual texture discrimination. *Spatial Vision* 4. 31-44.

[24] Heeger D. J. (1992) Normalization of cell responses in cat striate cortex. *Visual Neurosci.* **9**, 181-197.

[25] Haralick R. M. Shapiro L. G. (1992) *Computer and robot vision* Vol 1. Addison-Weslley Publishing.

[26] Hirsch J. A. and Gilbert C. D. (1991) Synaptic physiology of horizontal connections in the cat's visual cortex. *J. Neurosci.* **11**(6): 1800-9.

[27] Hubel D. H. and Wiesel T. N. (1962) Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol.* **160**, 106-154.

[28] Julesz, B. (1962) Visual pattern discrimination *IRE Transactions on Information theory IT-8* 84-92.

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

[30] Kapadia, M. K., Ito, M. , Gilbert, C. D., and Westheimer G. (1995) Improvement in visual sensitivity by changes in local context: parallel studies in human observers and in V1 of alert monkeys. *Neuron*. **15**(4), 843-56.

[31] Karni A, Sagi D. (1991) Where practice makes perfect in texture discrimination: evidence for primary visual cortex plasticity. *Proc. Natl. Acad. Sci. USA* 88 (11): 4977.

[32] Kastner S., Nothdurft H-C., Pigarev I.N. (1997) Neuronal correlates of pop-out in cat striate cortex. *Vision Res.* **37**(4), 371-76.

[33] Knierim J.J. and van Essen D. C. (1992) Neuronal responses to static texture patterns ion area V1 of the alert macaque monkeys. *J. Neurophysiol.* **67**, 961-980.

[34] Kovacs I. and Julesz B. (1993) A closed curve is much more than an incomplete one: effect of closure in figure-ground segmentation. *Proc Natl Acad Sci USA*. 15; 90(16): 7495-7.

[35] Lamme V.A. (1995) The neurophysiology of figure-ground segregation in primary visual cortex. *Journal of Neuroscience* **15**(2), 1605-15.

[36] Lamme V. A. F., Zipser K. and Spekreijse H. (1997) Figure-ground signals in V1 depend on consciousness and feedback from extra-striate areas *Soc. Neuroscience Abstract* 603.1.

[37] Levitt J. B. and Lund J. S. (1997) Contrast dependence of contextual effects in primate visual cortex. *Nature* **387**(6628), 73-6.

[38] Li, Zhaoping (1997) Primary cortical dynamics for visual grouping in *Theoretical aspects of neural computation* Wong, K.Y.M, King, I, and D-Y Yeung, (Eds), page 155-164. Springer-Verlag, Hong Kong.

[39] Li, Zhaoping. (1998a) A neural model of contour integration in the primary visual cortex. *Neural Computation* 10(4) p 903-940.

[40] Li, Zhaoping (1998b) Visual segmentation without classification: a proposed function for primary visual cortex. *Perception* vol. 27, supplement, p.45.

[41] Li, Zhaoping (1999) "A V1 model of pop out and asymmetry in visual search" to appear in *Advances in neural information process systems 11*, Kearns, M.S., Solla S.A., and Cohn D. A. Eds. MIT Press.

[42] Li, C.Y., and Li, W. (1994) Extensive integration field beyond the classical receptive field of cat's striate cortical neurons — classification and tuning properties. *Vision Res.* **34** (18), 2337-55.

[43] Livingstone M. S. and Hubel, D. H. (1984) Anatomy and physiology of a color system in the primate visual cortex. *J. Neurosci.* Vol. 4, No.1. 309-356.

[44] Malik J. and Perona, P. (1990) Preattentive texture discrimination with early vision mechanisms. *J. Opt. Soc. Am. A* **7**(5),923-932.

[45] Marr D. (1982) *Vision, A computational investigation into the human representation and processing of visual information*. Freeman.

[46] Merigan W. H. Mealey T.A., Maunsell J. H. (1993) Visual effects of lesions of cortical area V2 in macaques. *J. Neurosci* **13** 3180-91.

[47] Nothdurft H. C. (1985) Sensitivity for structure gradient in texture discrimination tasks. *Vis. Res.* Vol. 25, p 1957-68.

[48] Nothdurft H. C. (1991) Texture segmentation and pop-out from orientation contrast. *Vis. Res.* Vol. 31, p 1073-78.

[49] Nothdurft H. C. (1994) Common properties of visual segmentation. in *Higher-order processing in the visual system* Bock G. R., and Goode J. A. (Eds). (Wiley & Sons), pp245-268

[50] Nothdurft H-C., Gallant J.L. Van Essen D. C. (1998). Response modulation by texture surround in primate area V1: correlates of 'pop-out' under anesthesia. *Visual Neurosci.* in press.

[51] Nothdurft H-C. Different approaches to the coding of visual segmentation. in *Computational and psychophysical mechanisms of visual coding*, Eds. L. Harris and M. Kjenkius, Cambridge Unviersity Press, New York, 1997.

[52] Polat U. Mizobe K. Pettet M. Kasamatsu T. Norcia A. (1998) Collinear stimuli regulate visual responses depending on cell's contrast threshold. *Nature* vol. 391, p. 580-583.

[53] Polat U. and Sagi D. (1993) Lateral interactions between spatial channels: suppression and facilitation revealed by lateral masking experiments. *Vis. Res.* 33, p993-999.

[54] Polat U. and Sagi D. (1994) Spatial interactions in human vision: From near to far via experience-dependent cascades of connections. *Proc. Natl. Acad. Sci. USA* 91. p1206-1209.

[55] Rockland K. S. and Lund J. S. (1983) Intrinsic Laminar lattice connections in primate visual cortex. *J. Comp. Neurol.* **216**, 303-318

[56] Scialfa C. T. and Joffe K. M. (1995) Preferential processing of target features in texture segmentation. *Percept Psychophys* 57(8). p1201-8.

[57] Sillito, A. M. Grieve, K.L., Jones, H.E. Cudeiro, J. and Davis J. (1995) Visual cortical mechanisms detecting focal orientation discontinuities. *Nature* **378**(6556), 492-6.

[58] Sireteanu R. and Rieth C. (1992). Texture segregation in infants and children. *Behav. Brain Res.* 49. p. 133-139

[59] Somers, D. C., Todorov, E. V., Siapas, A. G. and Sur M. (1995) Vector-based integration of local and long-range information in visual cortex. *A.I. Memo. NO. 1556,* MIT.

[60] Stemmler M, Usher M, Niebur E (1995) Lateral interactions in primary visual cortex: a model bridging physiology and psychophysics. *Science* **269**, 1877-80.

[61] Treisman A. and Gormican S. (1988) Feature analysis in early vision: evidence for search asymmetries. *Psychological Rev.* **95**, 15-48.

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

[63] Ts'o D. and Gilbert C. (1988) The organization of chromatic and spatial interactions in the primate striate cortex. *J. Neurosci.* 8: 1712-27.

[64] Ullman, S (1994). Sequence seeking and counterstreams: A model for bidirectional information flow in the cortex. In C Koch and J Davis, editors, *Large-Scale Theories of the Cortex.* Cambridge, MA: MIT Press, 257-270.

[65] Weliky M., Kandler K., Fitzpatrick D. and Katz L. C. (1995) Patterns of excitation and inhibition evoked by horizontal connections in visual cortex share a common relationship to orientation columns. *Neurons* **15**, 541-552.

[66] White E.L. (1989) *Cortical circuits* Birkhauser.

[67] Wolfe J. M. and Friedman-Hill S. R. (1992) Visual search for oriented lines: The role of angular relation between targets and distractors. *Spatial Vision* Vol. 6., p 199-207.

[68] Wolfson S. and Landy M. S. (1995) Discrimination of orientation-defined texture edges. *Vis. Res.* Vol. 35, Nl. 20, p 2863-2877.

[69] Zipser, K. Lamme, V. A. and Schiller P. H. (1996) Contextual modulation in primary visual cortex. *J. Neurosci.* **16** (22), 7376-89.