

Identification of Salient Contours in Cluttered Images

Shih-Cheng Yen and Leif H. Finkel
Department of Bioengineering
University of Pennsylvania

Abstract

We present a model of contour extraction in which the perceptual salience of contours arises from long-range interactions between orientation-selective filters. Ullman [19], Zucker [22, 23] and colleagues have previously shown that salient contours may be extracted from noisy images by using a number of heuristic features. Our algorithm is based on cortical mechanisms, and simulations show close agreement with results from recent anatomical, physiological and psychophysical studies including recent results of Field *et al.* [3], Kovács *et al.* [10, 11, 12], and Kapadia *et al.*, [8]. The performance of the algorithm is demonstrated on a range of psychophysical stimuli and real images.

1. Introduction

Contour extraction has been the focus of many previous computational studies, yet it remains a difficult problem in practice. As pointed out by Sha'ashua and Ullman [19], the relative salience of the contours in an image suggests something about the cortical mechanisms used to extract them. Sha'ashua and Ullman defined salience by several heuristic features -- contour length, curvature, discontinuities, and gap sizes. Salient contours in noisy images were successfully identified by maximizing an objective function based on these features, over all possible contours. An alternative approach to contour extraction was developed by Zucker and colleagues [15, 22, 23] based on spline interpolation between oriented edges. Their model uses a two-stage process of identifying the contour and then precisely localizing its position. Guy and Medioni [7] have recently proposed an "extension field" which allows the processing of globally salient contours to be non-iteratively computed using local operators. These models perform well, however they are not addressed towards the underlying biological mechanisms.

Recent physiological and psychophysical "pop-out" experiments have provided insights into the mechanisms used by the visual system to identify salient contours [3, 8, 10, 11, 12, 16, 17, 18]. We present a biologically based model with long-range interconnections between linear filters which is able to account for these recent results, and

which successfully extracts salient contours from real images.

2. Model Architecture

Orientation-selective cells are represented by linearly separable G4 and H4 steerable filters [4]. The dominant orientation at each position is computed directly from basis responses using the method proposed in [4], and responses are squared to allow interactions between cells of opposite contrast polarity. Each cell receives weighted inputs from other oriented cells in its surround via long-range horizontal connections. The connection weights depend on position as well as orientation. Steerable filters allow these facilitatory and inhibitory interactions to be computed in an efficient manner.

2.1. Facilitation

The facilitation received by a postsynaptic cell of orientation θ is given by:

$$F(x, y, \theta) = \sum_{(i,j)} W(\theta, i, j, \theta) A(x+i, y+j, \theta) d_{ij}$$

where $A(x+i, y+j, \theta)$ is the activity of the presynaptic cell of orientation θ at position (i, j) and $W(\theta, i, j, \theta)$ is the connection weight from this presynaptic cell to the postsynaptic cell of orientation θ at (x, y) .

The sign, magnitude and time course of the synaptic interactions depend upon the position and orientation of the target cell, creating separate spatial zones of excitation and inhibition. The connection field is shown in Figure 1b. Excitatory connections are confined to two regions, one flaring out along the axis of orientation of the cell (co-axial), and another confined to a narrow zone extending orthogonally to the axis of orientation (trans-axial). The co-axial connection pattern is similar to the "association field" proposed by Field *et al.* [3] and the "extension field" of Guy and Medioni [7], and is generated by a modification of Parent and Zucker's [15] "co-circularity" connection equation. (The "co-circular" connection pattern serves as a model for all the possible smooth curves that could pass through both "A" and "B" and is not a circle or curvature detector.) Given the orientation, θ , of the postsynaptic

cell, the “preferred” orientation, θ , at the position (i,j) of the presynaptic cell is specified by:

$$\theta(i, j) = 2 \tan^{-1} \frac{j}{i} - \theta_0$$

where θ_0 is the orientation of the postsynaptic cell, and i, j are positions relative to the postsynaptic cell (see Figure 1a). The connection weights peak at θ_0 and fall off as a Gaussian with half-width at half-height, a :

$$B(\theta, i, j) = G(\theta - \theta_0, a)$$

Based on psychophysical measurements by Polat and Sagi [17, 18] as well as physiological studies by Kapadia *et al.*, [8] the connection weights also fall off as a Gaussian function of distance,

$$D(i, j) = G(\sqrt{i^2 + j^2}, a_d)$$

where a_d represents the half-width at half-height of the Gaussian distribution.

Connection fan-out is limited to low curvature deviations from the orientation axis and to a narrow region extending orthogonal to the cell's orientation axis. These constraints may be expressed as:

$$W(i, j) = \begin{cases} 1, & \text{if } \left| \tan^{-1} \frac{j}{i} - \theta_0 \right| \leq \theta_c \\ 1, & \text{if } \left| \tan^{-1} \frac{j}{i} - \theta_0 - \frac{\pi}{2} \right| \leq \theta_t \\ 0, & \text{otherwise.} \end{cases}$$

where θ_c represents the maximum angular deviation of the co-axial connections and θ_t represents the maximum angular deviation of the trans-axial connections.

Thus the facilitation for a cell of orientation θ , located at position (x,y) , can be represented by:

$$F(x, y) = \sum_{(i,j) \in N} W(i, j) D(i, j) B(\theta, i, j)$$

Figure 1b illustrates the connection pattern for a horizontally oriented cell.

2.2. Dynamic Threshold

Long-distance connections provide support for incorporation of local oriented elements into a contour. The range over which such support extends depends upon the signal/noise ratio in the image. In psychophysical experiments, Kovacs *et al.* [12] have shown that contour detectability depends upon the ratio of the spacing between contour elements versus the spacing between “distractor” elements (see section 3.2).

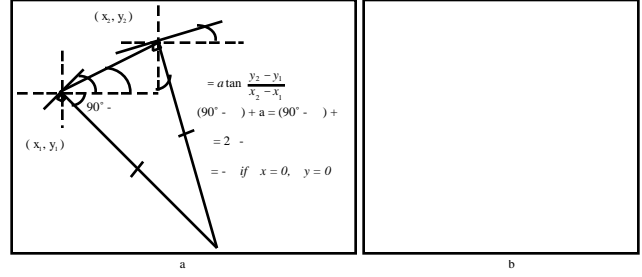


Figure 1. a) Co-circularity constraint. b) Connectivity pattern of a horizontally oriented cell. Length of line indicates connection strength.

In the model, this signal/noise effect is carried out by a set of non-oriented cells whose inputs are weighted by distance but not orientation. The activity of these cells provides a measure of the density of background elements in the image. The non-oriented cells determine the facilitation threshold for the orientation-selective cells. Long-distance inputs to orientation-selective cell must exceed this threshold to be facilitatory. The threshold, θ , depends upon the magnitude of the input, I , as given by

$$\theta = e^{-I}$$

where $\theta = 9$ in all simulations. The value of θ determines the distance over which facilitation occurs. This choice of θ corresponds to facilitation out to a distance of 1.5 times the average separation, s , of the background elements. Inputs from nearby cells which do not have the right orientation will not be facilitatory since connection weights fall-off sharply with orientation (e.g. $\theta = 20^\circ$).

2.3. Inhibition

The first stage of the model identifies those cells that are receiving long-range facilitated inputs in addition to direct visual (thalamic) input. The next stage of the model incorporates a longer-latency inhibition that originates outside the facilitatory zones, as suggested by the physiological results of Nelson and Frost [13, 14] and Kapadia *et al.* [8]. The magnitude of this inhibition is set such that it is strong enough to suppress cells with weak support but not sufficient to suppress cells with strong facilitation [13, 14, 8]. Cells mutually inhibit each other such that the facilitation provided by the presynaptic cell is effectively divided by the number of postsynaptic cells. Each of these postsynaptic cells may also be facilitated by additional cells. Cells are only considered facilitated if their total support exceeds a fixed threshold (in all simulations reported, set to 0.5).

2.4. Synchronization and Saliency

Facilitatory and inhibitory processes operate in parallel to identify units lying on contours across the image. However, in order to discriminate separate contours from each other, the responding units must be bound into coherent groups based on the information present in the scene. As Singer and Gray [20] have observed, binding must be flexible since relationships will vary for different visual scenes. One possible mechanism underlying this process may be the synchronization of cell firing [20]. Gray and McCormick [6] have recently shown that the same supra-granular simple cells which send off horizontal connections are capable of bursting, and thus may be involved in binding processes.

In the model, facilitated cells are assumed to enter a “bursting” mode. Synchronization of these cells is modeled using coupled neural oscillators. (We use neural oscillators only as a simple functional means of computing synchronization and make no assumption about the functional role of neural oscillations in cortex. See [1, 5].) Oscillators are coupled to other oscillators with which they have strong, reciprocal, facilitated connections, and which are also within a threshold distance, r . This threshold is determined by the background separation (in all simulations $r = 2.8s$). The oscillators are initialized with random phases from $0^\circ - 360^\circ$. The oscillators synchronize using a simple phase averaging rule:

$$\theta_i(t) = \frac{1}{N} \sum_j \theta_j(t-1)$$

where θ_i represents the phase of the N oscillators, and the average is computed using a vector representation to avoid the discontinuity at 360° .

The oscillators synchronize iteratively with synchronization defined using the following condition:

$$\left| \theta_i(t) - \theta_j(t) \right| < \theta, \quad i, j \in C, \quad t < t_{\max}$$

where C represents all the coupled oscillators on the same contour, θ represents the maximum phase difference between oscillators, and t_{\max} represents the maximum number of time steps the oscillators are allowed to synchronize. A contour is thus represented by a synchronized chain. The saliency of the contour is given by the sum of the activities of all the synchronized elements on the contour, S_C . The contour with the highest saliency is chosen as the output of the network. This allows us to compare the output of the model to psychophysical results on contour extraction.

3. Results

Simulations of the model can be tested against reported psychophysical experiments. All simulations were conducted with the same set of parameters

3.1. Extraction of Salient Contours

Using the same methods as described by Field *et al.* [3], we generated stimuli consisting of arrays of 256 oriented Gabor elements. Pairs of stimulus arrays were presented to the network: one array contained a contour composed of 12 Gabor elements, the other contained only randomly oriented elements. For each stimulus, the network determines the “saliency” of all contours, and selects the contour with the highest saliency. And for the two stimuli in each pair, the stimulus containing the contour with the higher saliency is chosen. Network performance was measured by computing the percentage of correct detection in this 2AFC paradigm, as shown in Figure 4. As the difference in angle, θ , between neighboring contour elements was increased, the performance of the network was measured under the following conditions: a) elements aligned parallel to the contour, b) elements aligned orthogonal to the contour, c) elements randomly rotated ($\theta = 30^\circ$) away from the contour, and d) average separation of all the elements increased from 0.5° (32 pixels in the model) to 0.9° (64 pixels in the model). 500 simulations were run at each data point. Results are shown in Figure 2. When the elements are aligned, the performance of both the network and human subjects decreases with increasing θ (Figure 2a). In the model, as θ increases, there is an increased likelihood that the connections between consecutive elements will fall outside the facilitatory zone. This reduces the number of elements bound together on the target contour and causes a decrease in saliency. This accounts for the much sharper drop in performance of the model at 60° as compared to data from human psychophysics. A better approximation to the data would result from a gradual decrease in the connection strengths as positions deviate from the orientation axis of the postsynaptic cell. The model is also able to extract contours when the elements are aligned orthogonal to the contour. This ability is modulated by the trans-axial connections. The fan-out of the facilitatory connections is narrower ($\pm 15^\circ$) than that of the co-axial connections and thus the performance of the network falls off more rapidly with increasing θ , reaching chance performance at $\pm 30^\circ$. Again, the network’s performance is comparable to the reported psychophysical data (Figure 2b).

When the orientation of the elements are randomly offset with respect to the path of the contour, they deviate from the “preferred” orientations at their positions and thus

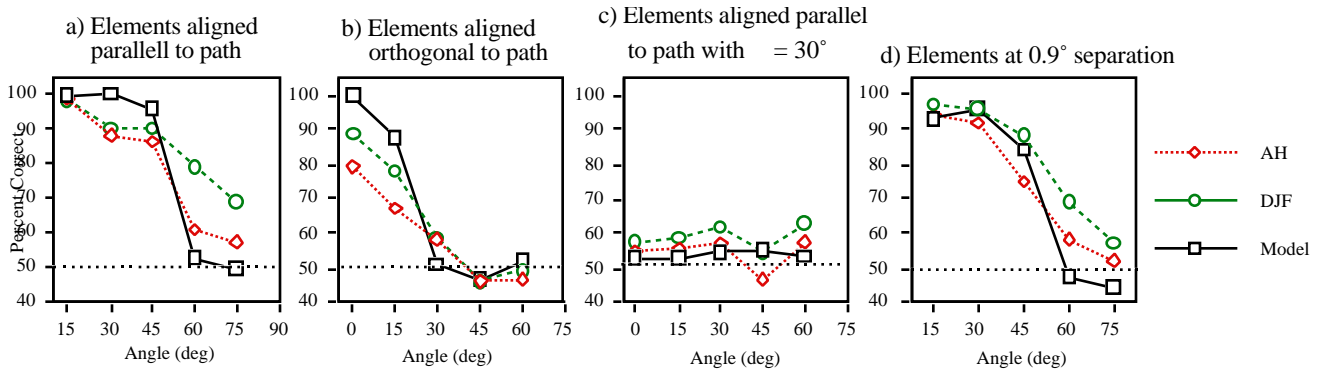


Figure 2. Results from the simulations of the experiments of Field *et al.* [3]. Stimuli consisted of 256 randomly oriented Gabor patches. Contours were formed by aligning 12 elements either collinear or orthogonal to the orientation axis of the Gabor elements. Each data point represents the performance for 500 simulations. The simulation results are compared to the psychophysical data from the two subjects (AH, DJF) in [3].

reduce the strength of the connections. This increases the likelihood that these inputs may fall below the threshold imposed by the background noise. This leads to “breakage” in the chains and leads to a decrease in the saliency of the contours. The model demonstrates the same qualitative behavior as the data for $\theta = 30^\circ$ (Figure 2c).

As the average separation between elements is increased, the degree of facilitation is decreased. However, since inputs from background elements are also decreasing in strength, the signal-to-noise ratio is not altered. This allows the model to continue to detect the contour over a wide range of absolute separation distances between the elements. This behavior agrees well with the psychophysical data (Figure 2d).

3.2. Effects of Contour Closure

In a series of experiments using stimuli similar to Field *et al.* [3], Kovács and Julesz [10] found that closed contours are much more salient than open contours. More recently, Kovács *et al.* [12], showed that elements on a circular contour could be separated up to 1.5 times further than the average separation of the background elements and still remain salient. While the maximum separation for elements on an open contour (θ_o) was not explicitly determined, their initial results suggest that it would be significantly less than that for closed contours (θ_c). In addition, they showed that when elements spaced at θ_o are added to a short “jagged” (open) contour, the saliency of the contour increases monotonically but when elements spaced at θ_c are added, the saliency does not change until the last element is added and the contour becomes closed. In fact, at θ_c , the contour is not salient until it is closed, at which point it suddenly “pops-out” (see Figure 3c). This finding places a strong constraint on the computation of saliency in visual perception.

As in [10], we generated stimulus arrays containing 2025 elements. Contours were made up of 24 elements. Again the network is presented with two stimuli, one containing a contour and the other made up of all randomly oriented elements. The network picks the stimulus containing the synchronized contour with the highest saliency. 500 trials were simulated for each data point. In separate blocks, the contour elements were placed at increasingly greater separations, while the background elements remained at the same separation (set to 25 pixels in our simulations). Since the strength of the coupling between oscillators is dependent on the signal-to-noise ratio of the facilitatory weights, the synchronization falls apart as the facilitatory inputs from the contour elements decrease in magnitude. However, it has been shown that oscillators on a closed chain can synchronize much better than those on an open chain, as a result of the boundary conditions [2, 9, 21]. Our simulations also show that elements on a closed chain are able to synchronize at higher separations compared to the open chains. The threshold ratio of the contour separation to the background separation ($\theta_c = \text{noise separation} / \text{contour separation}$), defined at 75% accuracy, for open (θ_o) and closed (θ_c) contours were determined, as shown in Figure 3a. The results show θ_o to be about 0.9 while θ_c is about 0.6 -- a ratio of 1.5.

We also examined the changes in saliency for open and closed contours. Stimuli containing 2025 elements were generated with contour elements spaced at θ_o and θ_c . The background separation was kept constant for all stimuli. The performance of the network was measured as additional elements were added to an initial short contour of elements. The results are shown in Figure 3b. For open contours, each additional element synchronizes with the rest of the contour and adds to the saliency of the contour. This increases the probability that the stimulus containing the target contour will be picked by the network. For

closed contours, the initial number of elements are unable to synchronize as they are separated further than the threshold for the open contours. Since the target contour is unable to synchronize, the network’s comparison in the two-alternative forced choice paradigm is based on the saliency of other synchronized contours in the stimulus. Only when the last element is added and the contour is closed, is the network able to select the target as the contour with the highest saliency and thus show greatly increased detectability. This matches the results of Kovács and Julesz [10], where the saliency of the closed curve does not change significantly until the last element is added, thus “closing” the chain and causing it to “pop-out”.

3.3. Real Images

A more stringent test of the model’s capabilities is the ability to extract perceptually salient contours in real images. Figure 5 and 6 show results for a typical image. The original grayscale image, the output of the steerable filters, and the output of the model are shown in Figure 5a,b,c and Figure 6a,b,c respectively. The network is able to extract some of the more salient contours and ignore other high contrast edges detected by the steerable filters. Figure 5 illustrates how camouflage attempts to re-order the saliency of contours. Most of the edges of the plane are almost invisible due to the low contrast at the edges, and the similarity in texture to the background. The highest contrast edges correspond to the camouflage markings on the plane. Nonetheless, the network extracts the plane edges and chooses them as most salient due to their length and straightness. Figure 6 represents the other extreme where high-contrast edges are found over the whole image. The model is still able to extract some of the more salient contours like the ones representing the bridge, as well as the distinctive circular contours on buildings. Both simulations used filters at only one spatial scale and could be improved through interactions across multiple spatial frequencies. Nevertheless, the model shows promise for automated image processing applications.

4. Discussion

Experimental results from physiology and psychophysics have suggested that salient contours may be extracted through long-range interactions between oriented cells. We have incorporated these properties into a network of oriented filters and shown that a model with excitatory and inhibitory interactions, coupled with signal-to-noise constraints, is able to account for a number of experimental results. The network is also able to extract contours in a variety of real images and endow them with saliency. The network has applications in terms of identifying

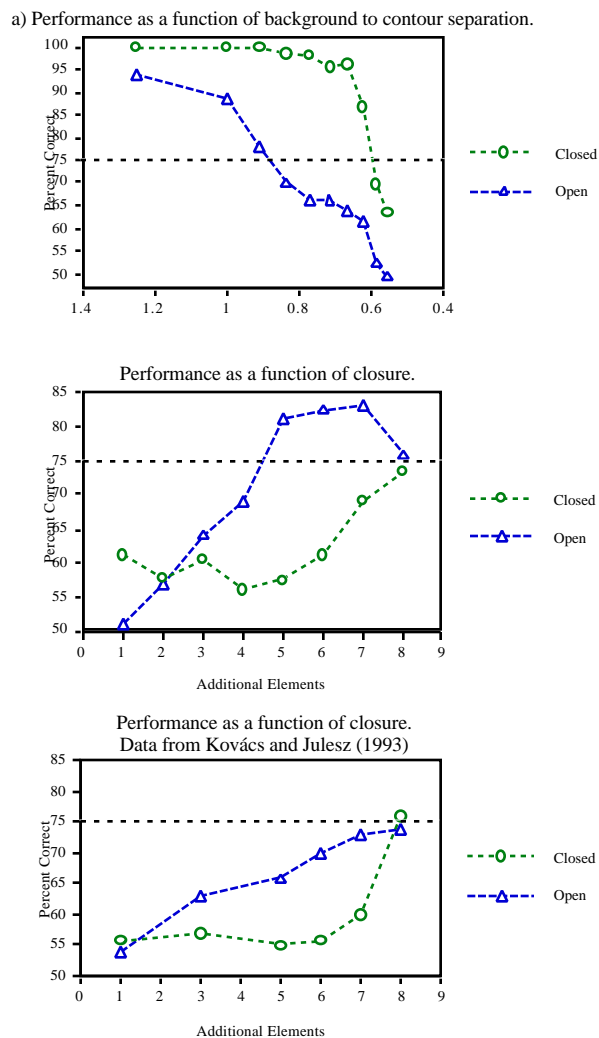


Figure 3. Simulation of the experiments of Kovács and Julesz [10]. Stimuli consisted of 2025 randomly oriented elements. Contours were made up of 24 elements. Each data point represents results from 500 simulations. a) Results show the difference in for open and closed contours. b) Changes in saliency with additional elements was computed for open and closed contours. Open contours monotonically increase in saliency with additional elements but closed contours become more salient only when the contour is closed or almost closed. Open contours started with 7 elements while closed contours started with 17 elements. c) The data from [10] is re-plotted for comparison.

“interesting” features in automatic image processing applications. The facilitatory architecture can be extended to multiple spatial frequencies, thus allowing a natural

mechanism to implement a hierarchical processing system with interactions between different spatial frequencies.

Acknowledgments

Supported by the Office of Naval Research (N00014-93-1-0681) and The Whitaker Foundation.

References

- [1] W. Bair, C. Koch, W. Newsome, and K. Britten. Power spectrum analysis of bursting cells in area MT in the behaving monkey. *Journal of Neuroscience*, 14:2870-2892, 1994.
- [2] G. B. Ermentrout. The behavior of rings of coupled oscillators. *Journal of Mathematical Biology*, 23:55-74, 1985.
- [3] D. J. Field, A. Hayes, and R. F. Hess. Contour integration by the human visual system: Evidence for a local "Association Field". *Vision Research*, 33:173-193, 1993.
- [4] W. T. Freeman and E. H. Adelson. The design and use of steerable filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13:891-906, 1991.
- [5] G. M. Ghose and R. D. Freeman. Oscillatory discharge in the visual system: Does it have a functional role? *Journal of Neurophysiology*, 68:1558-1574, 1992.
- [6] C. M. Gray and D. A. McCormick. Chattering cells -- superficial pyramidal neurons contributing to the generation of synchronous oscillations in the visual-cortex. *Science*, 274:109-113, 1996.
- [7] G. Guy and G. Medioni. Inferring global perceptual contours from local features. *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 786-787, 1993.
- [8] M. K. Kapadia, M. Ito, C. D. Gilbert, and G. Westheimer. Improvement in visual sensitivity by changes in local context: Parallel studies in human observers and in V1 of alert monkeys. *Neuron*, 15:843-856, 1995.
- [9] N. Kopell and G. B. Ermentrout. Symmetry and phaselocking in chains of weakly coupled oscillators. *Communications on Pure and Applied Mathematics*, 39:623-660, 1986.
- [10] I. Kovács and B. Julesz. A closed curve is much more than an incomplete one: Effect of closure in figure-ground segmentation. *Proceedings of National Academy of Sciences, USA*, 90:7495-7497, 1993.
- [11] I. Kovács and B. Julesz. Perceptual sensitivity maps within globally defined visual shapes. *Nature*, 370:644-646, 1994.
- [12] I. Kovács, U. Polat, and A. M. Norcia. Breakdown of binding mechanisms in amblyopia. *Investigative Ophthalmology & Visual Science*, 37:3078, 1996.
- [13] Nelson, J. I. & Frost, B. J. (1978). Orientation-selective inhibition from beyond the classic visual receptive field. *Brain Research*, 139, 359-365.
- [14] Nelson, J. I. & Frost, B. J. (1985). Intracortical facilitation among co-oriented, co-axially aligned simple cells in cat striate cortex. *Experimental Brain Research*, 61, 54-61.
- [15] P. Parent and S. W. Zucker. Trace inference, curvature consistency, and curve detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11:823-839, 1989.
- [16] M. W. Pettet, S. P. McKee, and N. M. Grzywacz. Smoothness constrains long-range interactions mediating contour-detection. *Investigative Ophthalmology and Visual Science*, 37:4368, 1996.
- [17] U. Polat and D. Sagi. Lateral interactions between spatial channels: Suppression and facilitation revealed by lateral masking experiments. *Vision Research*, 33:993-999, 1993.
- [18] U. Polat and D. Sagi. The architecture of perceptual spatial interactions. *Vision Research*, 34:73-78, 1994.
- [19] A. Sha'ashua and S. Ullman. Structural saliency: the detection of globally salient structures using a locally connected network. *Proceedings of the second international conference on computer vision*, 321-327, 1988.
- [20] W. Singer and C. M. Gray. Visual feature integration and the temporal correlation hypothesis. *Annual Review of Neuroscience*, 18:555-586, 1995.
- [21] D. Somers and N. Kopell. Rapid synchronization through fast threshold modulation. *Biological Cybernetics*, 68:393-407, 1993.
- [22] S. W. Zucker, C. David, A. Dobbins, and L. Iverson. The organization of curve detection: coarse tangent fields and fine spline covering. *International Conference on Computer Vision*, 2:1-10, 1989.
- [23] S. W. Zucker, A. Dobbins, and L. Iverson. Two stages of curve detection suggest two styles of visual computation. *Neural Computation*, 1:68-81, 1989.

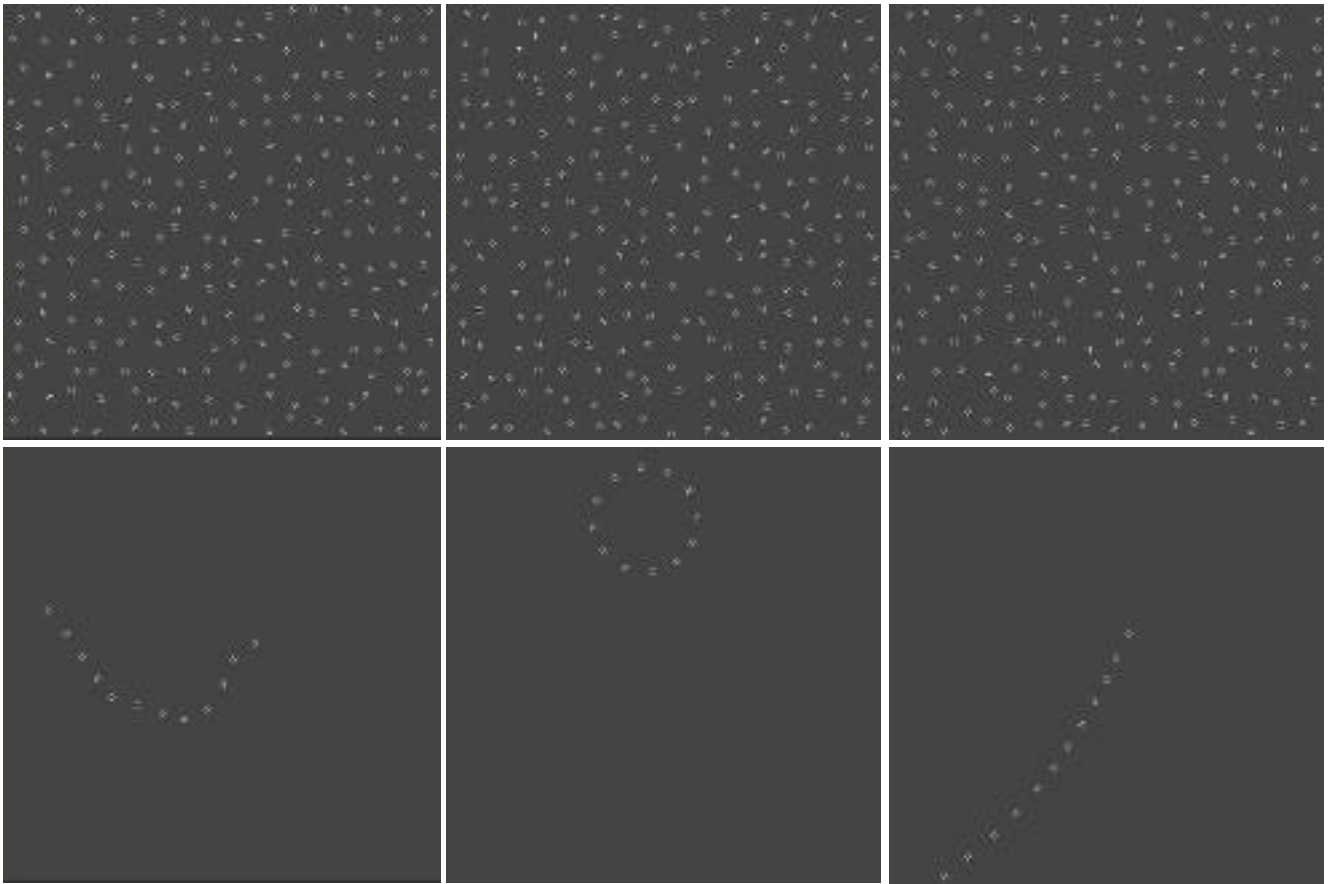


Figure 4. Salient contours extracted from sample input stimulus containing an open contour (left), a closed contour (middle), and a line with elements oriented orthogonal to the contour (right).

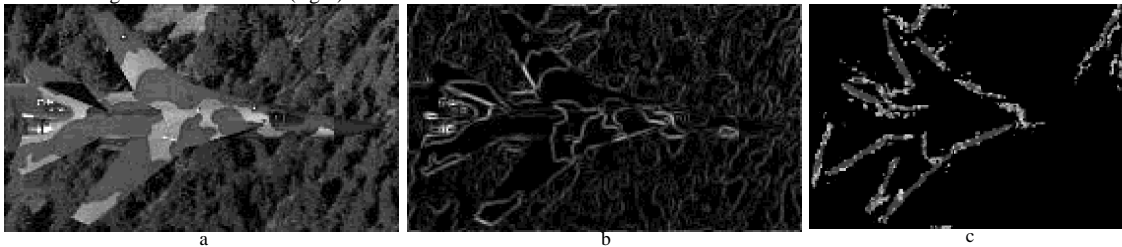


Figure 5. a) Plane Image. b) Steerable filter response. c) Result of model showing the most salient contours.

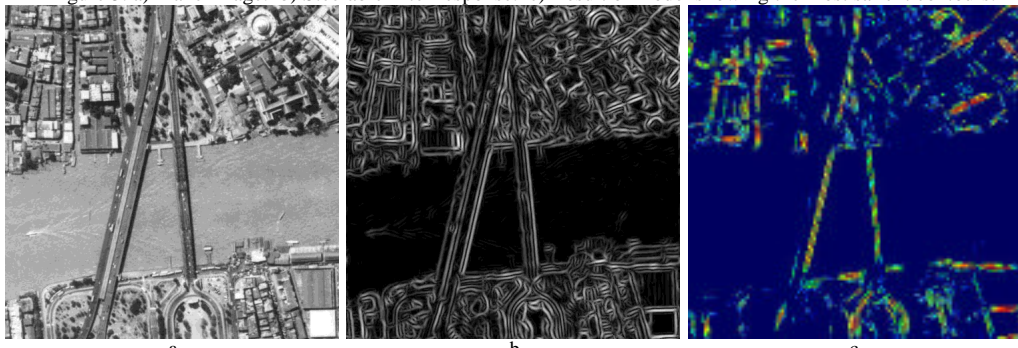


Figure 6 a) Satellite image of Bangkok. b) Steerable filter response. c) Salient contours extracted from the image. The model included filters at only one spatial frequency.