

# A Comparison of Measures for Detecting Natural Shapes in Cluttered Backgrounds

Lance R. Williams

Karvel K. Thornber

Dept. of Computer Science  
University of New Mexico  
Albuquerque, NM 87131

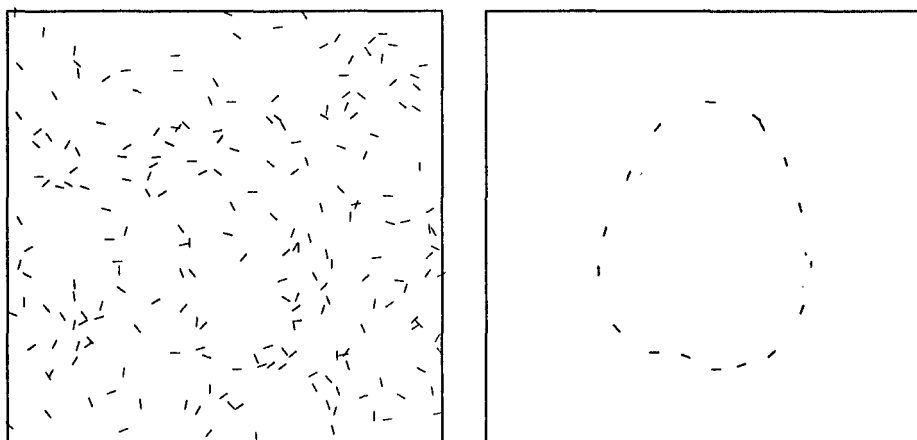
NEC Research Institute  
4 Independence Way  
Princeton, NJ 08540

**Abstract.** We propose a new measure of perceptual saliency and quantitatively compare its ability to detect natural shapes in cluttered backgrounds to five previously proposed measures. As defined in the new measure, the saliency of an edge is the fraction of closed random walks which contain that edge. The transition probability matrix defining the random walk between edges is based on a distribution of natural shapes modeled by a stochastic motion. Each of the saliency measures in our comparison is a function of a set of affinity values assigned to pairs of edges. Although the authors of each measure define the affinity between a pair of edges somewhat differently, all incorporate the Gestalt principles of good-continuation and proximity in some form. In order to make the comparison meaningful, we use a single definition of affinity and focus instead on the performance of the different functions for combining affinity values. The primary performance criterion is accuracy. We compute false-positive rates in classifying edges as signal or noise for a large set of test figures. In almost every case, the new measure significantly outperforms previous measures.

## 1 Introduction

The goal of segmentation is to partition a set of image measurements (e.g., edges) into equivalence classes corresponding to distinct objects. In this paper, we consider a somewhat simpler grouping problem which (following [11]) we call the *saliency problem*. The goal of the saliency problem is to assign a value to each edge which is correlated with whether that edge belongs to a shape or is background noise (see Figure 1). Given the distribution of saliency values, it is then often possible to choose a threshold which will segment the edges into shape and noise classes.

Each of the saliency measures proposed in the literature is a function of a set of affinity values assigned to pairs of edges. Although the authors of each measure define the affinity between a pair of edges somewhat differently, all incorporate the Gestalt principles of good-continuation and proximity in some form. A *saliency function* maps the set of affinities between all pairs of oriented or directed edges (i.e., the *affinity matrix*) to a *saliency vector*. In this paper, we have chosen to compare the definitions of saliency—not affinity. The differences in the authors' definitions of affinity prevents a direct comparison since



**Fig. 1.** Left: Seventeen edges from the boundary of a pear in a background of two-hundred edges of random position and orientation. Right: The most salient edges (where saliency is defined as the fraction of closed random walks which contain that edge.)

each requires its own set of parameters. The choice was either to 1) optimize the performance of each measure over its required parameters and compare the different measures with each using its optimal parameter setting; or 2) replace the individual affinity functions with a single function and compare performance using a single parameter setting. Apart from requiring an impractical amount of work, the first approach has the disadvantage of confounding the definitions of affinity and saliency so that the relative merits of each are difficult to disentangle. The shortcoming of the second approach (which is the one we adopted) is that while providing the best comparison of the saliency functions, it says nothing about the relative merits of the affinity functions. Although unlikely, it also ignores possible dependencies between the specific affinity and saliency functions used in a given measure.

The affinity functions can be divided into three classes. Functions in the first class are based on co-circularity[3, 5, 14]. The disadvantage of these functions is that they are non-generic—a circle does not have sufficient degrees of freedom to smoothly join two arbitrary positions and orientations in the plane. They are also difficult to motivate using arguments based on the statistics of natural shapes. Functions in the second class are based on curves of least energy[4, 11]. The affinity between two directed edges is inversely related to the energy,  $\int_T ds (\alpha \kappa^2(s) + \beta)$ , in the curve of least energy joining the two edges. Functions in the third class are based on an explicit prior probability distribution of natural shapes modeled by a stochastic motion[7, 13, 16]. A particle travels with constant speed in the direction  $\theta(t)$ . Change in direction is a normally distributed random variable with zero mean. Consequently,  $\theta(t)$  is a Brownian motion. The variance

of the random variable reflects the prior expectation of smoothness. In addition, a constant fraction of particles decay per unit time. The half-life reflects the prior expectation of shortness. The affinity between two edges,  $i$  and  $j$ , is defined as the sum of the probabilities over all paths joining the two edges, i.e.,  $P'(j|i)$ . Curves of least energy and stochastic motions are closely related. In fact, it is possible to show that the energy of the curve of least energy is a linear function of the log-likelihood of the maximum likelihood stochastic motion [7, 16]. It follows that the function  $P'(j|i)$  behaves very similarly to  $\exp[-\int_{\Gamma} dt (\alpha\kappa^2(t) + \beta)]$  when  $\Gamma$  is a curve of least energy. This is because the probability associated with  $\Gamma$  (and curves of similar shape) dominates the probabilities summed over all paths.

To facilitate the exposition, we will introduce a single nomenclature for describing all of the saliency measures. One of the major differences between the measures is whether they are formulated using orientations or directions. By direction we mean an angular quantity with a unique value over its  $2\pi$  range. By orientation we mean an angular quantity where a direction, e.g.,  $\theta$ , and its opposite direction, e.g.,  $\theta + \pi$  are identified with each other. Orientations assume values in the range zero to  $\pi$ . We will use  $\mathbf{x}$  to represent a vector of values associated with edge directions and  $\mathbf{y}$  to represent a vector of values associated with edge orientations. If a stimulus contains  $n$  edge segments then the saliency vector,  $\mathbf{x}$ , has  $2n$  components while the saliency vector,  $\mathbf{y}$ , has  $n$  components. The saliency vectors  $\mathbf{x}$  and  $\bar{\mathbf{x}}$  are identical except for a permutation which exchanges opposite directions. For example, if  $x_i$  represents the saliency associated with some edge with direction,  $\theta$ , then  $\bar{x}_i$  represents the saliency of the same edge, but in the opposite direction,  $\theta + \pi$ .

All of the saliency measures in our comparison associate an affinity value with a pair of oriented or directed edges. We will use the  $n \times n$  matrix  $A$  to represent the affinity values between all pairs of oriented edges and the  $2n \times 2n$  matrix  $P$  to represent the affinity values between all pairs of directed edges. An important distinction between saliency measures based on orientation and those based on direction involves the symmetry (or non-symmetry) of the affinity matrices. While the affinity oriented edge  $i$  has for  $j$  equals the affinity that oriented edge  $j$  has for  $i$ , this does not (generally) hold for directed edges. Basically,  $A_{ij} = A_{ji}$  (i.e.,  $A = A^T$ ) but in general,  $P_{ij} \neq P_{ji}$ . For reasons we will describe later (see Figure 4), this difference is critical in understanding the relative performance of the various measures.

Although not symmetric,  $P$  exhibits another kind of symmetry. If we use the subscript  $\bar{i}$  to denote the opposite direction to  $i$  then  $P_{ij} = P_{\bar{j}\bar{i}}$ . This is termed *time-reversal* symmetry. For the purposes of our comparison, we will define  $A_{ij}$  to be  $\max(P_{ij}, P_{\bar{i}\bar{j}}, P_{i\bar{j}}, P_{\bar{i}j})$ , that is, the affinity between two orientations is defined to be the maximum of the affinities among all combinations of directions.

## 2 Saliency Measures

In the following section, we provide short synopses of the saliency measures used in the comparison.

### 2.1 Shashua and Ullman (SU)

Shashua and Ullman[11] were the first to use the term saliency in the sense that it is being used in this paper. Building on earlier work by Montanari[6], they described a saliency measure which could be computed by a local parallel network.<sup>1</sup> Using our nomenclature, the saliency of a network element (one for each position and direction) at time  $t+1$  is related to the saliencies of neighboring elements at time  $t$  by the following update equation:

$$x_i^{(t+1)} = 1 + \max_j P_{ij} x_j^{(t)}$$

Recently, Alter and Basri[1] have done an extensive analysis of Shashua and Ullman's method and give expressions for the saliency measure for the case of continuous curves. The saliency of a directed edge  $i$  equals the maximum of the saliencies of all continuous curves,  $\Gamma$ , which begin at that edge:

$$\Phi(i) = \max_{\Gamma \in \mathcal{C}(i)} \Phi(\Gamma)$$

The saliency of a continuous curve,  $\Gamma$ , is given by the following expression:

$$\Phi(\Gamma) = \int_{s_1}^{s_n} ds \sigma(s) \cdot \rho \int_{s_1}^s dt (1 - \sigma(t)) \cdot e^{-\int_{s_1}^s dt \kappa^2(t)}$$

where  $\sigma(\cdot)$  is an indicator function which equals one where the curve lies on an edge element (and equals zero elsewhere),  $\rho$  is a parameter in the interval  $[0, 1]$  which controls the rate of convergence, and  $\kappa^2(\cdot)$  is the square of the curvature. The overall effect is that the Shashua-Ullman measure favors long smooth curves containing only a few short gaps. For the special case of a curve threading a sequence of  $n$  edges of negligible length, we have the following simplification:

$$\Phi(\Gamma) = \sum_{i=1}^n e^{-\int_{s_1}^{s_i} dt (\kappa^2(t) - \ln \rho)}$$

We observe that the first two terms of this series will dominate all subsequent terms unless the radius of curvature is large and the curve is densely sampled. Consequently, for visual patterns consisting of a sparsely sampled curve in a background of noise, the Shashua and Ullman measure becomes local and greedy. We will see that this seriously limits its performance on such patterns in the presence of correlated noise.

<sup>1</sup> In our nomenclature, Montanari's update equation is  $x_i^{(t+1)} = \min(-\ln P_{ij} + x_j^{(t)})$ . After  $t$  time-steps,  $x_i^{(t)}$  equals the energy of the minimum energy curve of length  $t$  beginning at edge  $i$ . In general, this quantity will not converge to a finite value as  $t$  goes to infinity.

## 2.2 Héroult and Horaud (HH)

Héroult and Horaud[5] cast the problem of segmenting a set of oriented edges into figure and ground as a quadratic programming problem which is solved by simulated annealing. The objective function consists of two terms,  $-E_{saliency} - E_{constraint}$ :

$$\min -\frac{1}{2} \mathbf{y}^T H \mathbf{y} - b^T \mathbf{y} \text{ for } \mathbf{y} \in \{-1, +1\}^n$$

where  $H_{ij} = A_{ij} - \alpha$  and  $b_i = \sum_j (A_{ij} - \alpha)$ . The affinity function used by Héroult and Horaud is based on co-circularity, smoothness and proximity. Héroult and Horaud say only that  $\alpha$  is a parameter related to the signal-to-noise ratio but do not say how it is chosen or provide the value they used in their experiments. Experimentally, we have found that their method is very sensitive to the choice of this parameter. If  $\alpha$  is too large, the solution consists of all  $-1$ 's (i.e., all ground) while if it is too small it consists of all  $+1$ 's (i.e., all figure). Determining the proper value of  $\alpha$  makes the job of fairly comparing Héroult-Horaud with measures lacking a comparable parameter difficult. Therefore (for the comparison) we decided to maximize  $E_{saliency}$  over 0-1 solution vectors with exactly  $m$  components equal to 1:

$$\max \mathbf{y}^T A \mathbf{y} \text{ for } \mathbf{y} \in \{0, 1\}^n \text{ and } \mathbf{y}^T \mathbf{y} = m$$

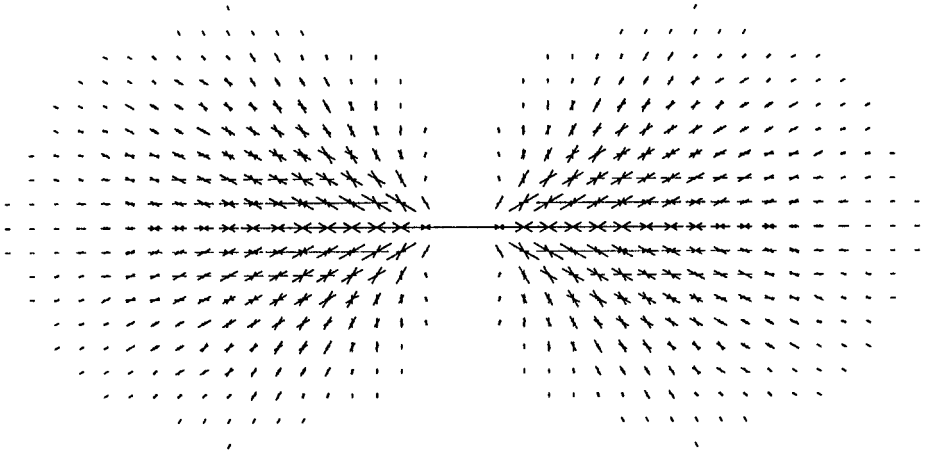
where  $m$  is the number of figure edges and  $n$  is the total number of edges. Although in a real application, we would generally not know the value of  $m$ , we do know this value for all of our test patterns. For this reason, the modified problem should provide a lower bound on the false-positive rate for the Héroult-Horaud measure.

## 2.3 Sarkar and Boyer (SB)

Sarkar and Boyer[10] describe a saliency measure and apply it to the problem of distinguishing developed and undeveloped land in aerial images. Although this is a somewhat different application than the one considered in this paper, the similarity between Sarkar and Boyer's computation and our own makes a comparison worthwhile. In addition to good-continuation and proximity, Sarkar and Boyer's affinity function incorporates pairwise measures useful for detecting clusters of buildings such as parallelism, perpendicularity and closure.<sup>2</sup> The affinity function we used in the comparison is the same one we used with the other methods (i.e., only a subset of the relations proposed by Sarkar and Boyer). Given an affinity matrix,  $A$ , Sarkar and Boyer propose that the saliency vector,  $\mathbf{y}$ , maximizes the Raleigh Quotient:

$$\frac{\mathbf{y}^T A \mathbf{y}}{\mathbf{y}^T \mathbf{y}}$$

<sup>2</sup> The closure between a pair of edges equals the closure of the group of edges containing that pair. Closed groups must be identified in advance.



**Fig. 2.** In the Yen and Finkel[17] saliency computation, the support for oriented edge  $i$  due to all other oriented edges  $j$  is given by the linear relaxation labeling step,  $y_i^{(t+1)} = \sum_j A_{j i} y_j^{(t)} / \sum_j \sum_k A_{j k} y_k^{(t)}$ . Because Yen and Finkel's intention was to model the visual cortex, the linear relaxation step is implemented as repeated convolution with a large kernel filter followed by normalization. The symmetry of the  $A$  matrix manifests itself in the plane as mirror image symmetry in the kernel. In general, this iteration will converge to the eigenvector associated with the largest positive real eigenvalue of  $A$ , i.e., it computes the same measure as Sarkar and Boyer[10].

When  $A$  is symmetric, the Raleigh Quotient is maximized by the eigenvector,  $\mathbf{y}$ , associated with the largest positive real eigenvalue of  $A$ :

$$\lambda \mathbf{y} = \mathbf{A} \mathbf{y}$$

This measure can also be optimized using the following recurrence equation (which has been independently proposed as a saliency computation by Yen and Finkel[17]):

$$y_i^{(t+1)} = \sum_j A_{i j} y_j^{(t)} / \sum_j \sum_k A_{j k} y_k^{(t)}$$

From linear algebra, we know that the vector  $\mathbf{y}$  will converge to the eigenvector associated with the largest positive real eigenvalue of  $A$ . Viewed this way, we see that  $A$  is being used as a linear relaxation labeling operator and that Sarkar and Boyer are solving a linear relaxation labeling problem as defined by Rosenfeld, Hummel and Zucker[9] (see Figure 2).

## 2.4 Guy and Medioni (GM)

Guy and Medioni[3] describe a saliency computation which involves the summation of vector voting patterns based on co-circularity and proximity. The distribution of votes which accumulate at a point in the plane is represented by its  $2 \times 2$  covariance matrix. The predominant orientation at a point is determined

by the eigenvector of the covariance matrix with largest eigenvalue. Neglecting the clever device of representing the vote distributions by their covariance matrices, it is possible to interpret Guy and Medioni's voting patterns as representing the correlation between orientations at different locations in the image plane. In our nomenclature, the saliency at an edge would be the sum of the voting patterns due to all other edges:

$$y_i = \sum_j A_{ij}$$

which is essentially one iteration of linear relaxation labeling using the operator  $A$  and a constant input vector.

## 2.5 Williams and Jacobs (WJ)

Williams and Jacobs[16] describe a method for computing a representation of illusory contours and occluded surface boundaries which they call a *stochastic completion field*. The magnitude of the stochastic completion field at  $(u, v, \phi)$  is the probability that a particle following a stochastic motion (representing the prior distribution of boundary completion shapes) will pass through  $(u, v, \phi)$  on a path joining two boundary fragments. Although not portrayed as a saliency measure, it is easy and natural to use this method to compute saliency. The saliency of an edge is defined to be the probability that a particle following a stochastic motion will pass through that edge on a path joining two others. The saliency vector is given by  $\mathbf{x} \bar{\mathbf{x}}$  where each component of  $\mathbf{x}$  is:

$$x_i = \sum_j P_{ij}$$

The value of  $x_i$  is the probability that a particle will reach directed edge  $i$  from some other edge  $j$ . The saliency of  $i$  is just  $x_i$  multiplied by  $\bar{x}_i$  (i.e., the probability that a particle will reach the same edge but with opposite direction). From time-reversal symmetry, we see that this equals the probability that a particle starting at any edge will pass through edge  $i$  and eventually reach another edge.<sup>3</sup>

## 3 A New Measure (WT)

We define the salience of an edge to be the fraction of closed random walks which contain that edge. It is important to distinguish between *random walks* and *stochastic motions*. By random walk, we mean a sequence of edges visited by a particle subject to the random process with transition probability matrix,  $P$ . By stochastic motion, we mean the path in the plane followed by a particle when  $\theta(t)$  is a Brownian motion. These two concepts are necessarily related, since the probability that a particle located at edge  $i$  at time-step  $t$  will be at edge  $j$  at

<sup>3</sup> It is also worth noting that the WJ measure can be computed very efficiently using a multi-resolution method. See [15].

time-step  $t + 1$  is defined to be the sum over the probabilities of all stochastic motions between  $i$  and  $j$ , i.e.,  $P_{ji} \equiv P'(j|i)$ . It follows that the distribution of random walks of length  $t + 1$  is related to the distribution of random walks of length  $t$  through multiplication by the matrix,  $P$ . If  $x_i^{(t)}$  represents the fraction of random walks of length  $t$  which end at directed edge  $i$ , then  $x_i^{(t+1)}$  (i.e., the fraction of length  $t + 1$  random walks), is given by the following recurrence equation (see Figure 3):

$$x_i^{(t+1)} = \sum_j P_{ij} x_j^{(t)} / \sum_j \sum_k P_{jk} x_k^{(t)}$$

The  $\sum_j \sum_k P_{jk} x_k^{(t)}$  term in the denominator is a normalization factor. Without the normalization after each step, the vector  $\mathbf{x}$  would quickly approach zero, because random walks of increasing length have decreasing probability. In the steady-state, this normalization factor equals  $\lambda$ :

$$\lambda \mathbf{x} = \mathbf{P} \mathbf{x}$$

where the eigenvector,  $\mathbf{x}$ , represents the fraction of random walks located at any given edge and the eigenvalue,  $\lambda$ , represents the ratio of the number of random walks which reach one more edge to the number which drift off or die in every step of the random process.<sup>4</sup> In the steady-state, the variation of the eigenvalue equals zero (i.e.,  $\delta\lambda/\delta\mathbf{x} = 0$ ) and the eigenvalue itself is given by the following equation:

$$\lambda = \frac{\bar{\mathbf{x}}^T \mathbf{P} \mathbf{x}}{\bar{\mathbf{x}}^T \mathbf{x}}$$

While  $\lambda$  can be set equal to other expressions, the significance of the above form is that it makes explicit the relationship between error in  $\mathbf{x}$  and error in  $\lambda$ . Specifically, it shows that if  $\mathbf{x}$  were in error by  $\delta\mathbf{x}$ , the calculated  $\lambda$  would be in error by only  $(\delta\mathbf{x})^2$ . Note that while this is a variational-principle for the eigenvalue, unlike the Raleigh Quotient, it is not a maximum-principle.<sup>5</sup>

It is possible to view the eigenvector,  $\mathbf{x}$ , as the distribution of random walks which survived in all past-times and (from time-reversability) to view  $\bar{\mathbf{x}}$  as the distribution of random walks which will survive in all future-times. By taking the product,  $\mathbf{x} \bar{\mathbf{x}}$ , we are constructing the distribution of random walks which survive in all past *and* future times. It follows that  $\mathbf{x} \bar{\mathbf{x}}$  represents the distribution of closed random walks through the edges.

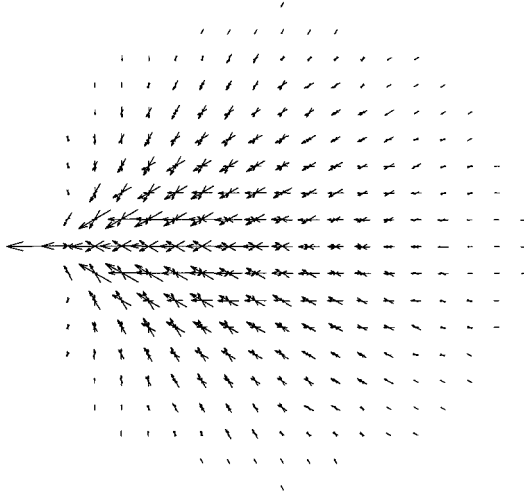
The use of directions (as opposed to orientations) is essential—even for non-directional stimuli.<sup>6</sup> Using only orientations, it is impossible to enforce tangent

<sup>4</sup> Unlike a Markov process,  $\lambda$  is usually very small—the great majority of particles never reach another edge. In a Markov process, the probabilities in every column of the transition matrix must sum to one. Consequently, the largest eigenvalue also equals one.

<sup>5</sup> In particular, there is no guarantee that a process which starts at a random vector and repeatedly applies the recurrence equation will converge to an eigenvector.

<sup>6</sup> The explicit representation of two directions is the motivation behind the design of the bipole cell in the boundary contour system of Grossberg and Mingolla[2]. It is also a feature of Ullman's model of illusory contour shape[14].





**Fig. 3.** In the saliency computation proposed in this paper, the support for directed edge  $i$  due to all other directed edges  $j$  is given by the linear relaxation labeling step,  $x_i^{(t+1)} = \sum_j P_{ji} x_j^{(t)} / \sum_j \sum_k P_{jk} x_k^{(t)}$ . Because  $P$  is not symmetric, this iteration is not guaranteed to converge. The saliency of edge  $i$  is given by the product of  $x_i$  and  $\bar{x}_i$  where  $\mathbf{x}$  is the eigenvector associated with the largest positive real eigenvalue of  $P$ . This quantity represents the relative number of closed random walks through edge  $i$ .

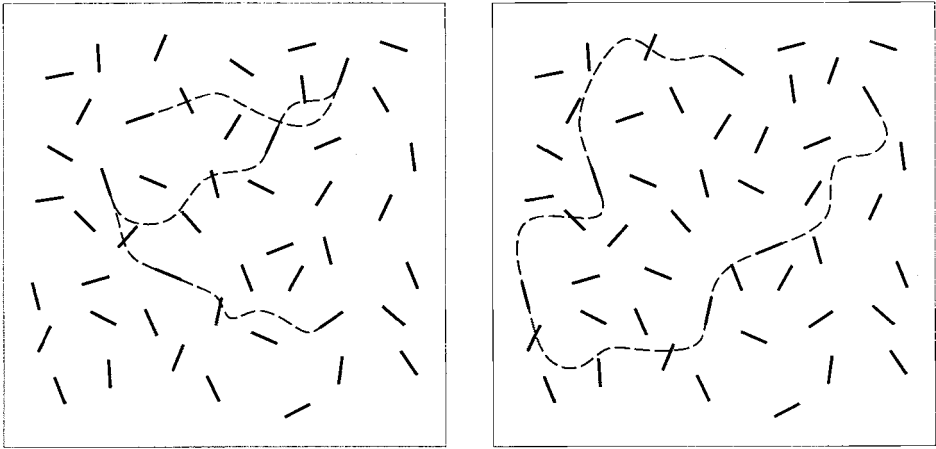
continuity at  $i$ . The consequence of not representing both directions is the presence of cusps (sudden reversals of direction) in the particle's paths (see Figure 4).

To develop some intuition for the meaning of the eigenvalue, it will be useful to consider an idealized situation. We know from linear algebra that the eigenvalues of  $P$  are solutions to the equation  $\det(P - \lambda I) = 0$ . Now, consider a closed path,  $\Gamma$ , threading  $m$  directed edges. The probability that a particle following this path will reach directed edge,  $\Gamma_{i \bmod m+1}$ , given that it is located at directed edge,  $\Gamma_i$ , equals  $P'(\Gamma_{i \bmod m+1} | \Gamma_i)$ . Assuming that the probability of a particle traveling from directed edge  $\Gamma_i$  to  $\Gamma_j$  when  $\Gamma_j$  does not immediately follow  $\Gamma_i$  on the closed path is negligible (i.e.,  $P_{ji} = P'(\Gamma_j | \Gamma_i)$  when  $j = i \bmod m + 1$  and  $P_{ji} = 0$  otherwise) then:

$$\lambda(\Gamma) = \left( \prod_{i=1}^m P'(\Gamma_{i \bmod m+1} | \Gamma_i) \right)^{\frac{1}{m}}$$

satisfies  $\det(P - \lambda I) = 0$ . This is the *geometric mean* of the transition probabilities in the closed path.<sup>7</sup> Normally long contours have very low probability:  $\prod_{i=1}^m P'(\Gamma_{i \bmod m+1} | \Gamma_i)$ . However, the properties of the geometric mean are such that smoothness and closure are favored and long contours suffer no penalty. It

<sup>7</sup> Equivalently, minus one times the logarithm of the eigenvalue equals the *average transition energy*:  $-\ln \lambda(\Gamma) = -\sum_{i=1}^m \ln P'(\Gamma_{i \bmod m+1} | \Gamma_i) / m$ .



**Fig. 4.** The explicit representation of two directions (i.e.,  $\mathbf{x}$  and  $\bar{\mathbf{x}}$ ) and the use of a *non-symmetric* linear relaxation labeling operator (i.e.,  $P$ ) is essential if the intermediate states of the relaxation labeling process (i.e.,  $x_i^{(t)}$ ,  $x_i^{(t+1)}$ , etc.) are to be interpreted as distributions of random walks which are continuous in tangent. Repeated application of a *symmetric* linear relaxation labeling operator (i.e.,  $A$ ) to a vector of saliencies associated with orientations (i.e.,  $\mathbf{y}$ ) yields distributions of random walks which can reverse direction at edge locations (left). After the initial iteration, this process tends to increase the salience of the noise edges as much as the signal edges. This explains why the performance of Guy and Medioni's saliency computation (essentially one iteration of linear relaxation labeling using the operator  $A$  and a constant input vector) is superior to that of Sarkar and Boyer[10]. It also explains why direction-based measures (e.g., WJ and WT) outperform both.

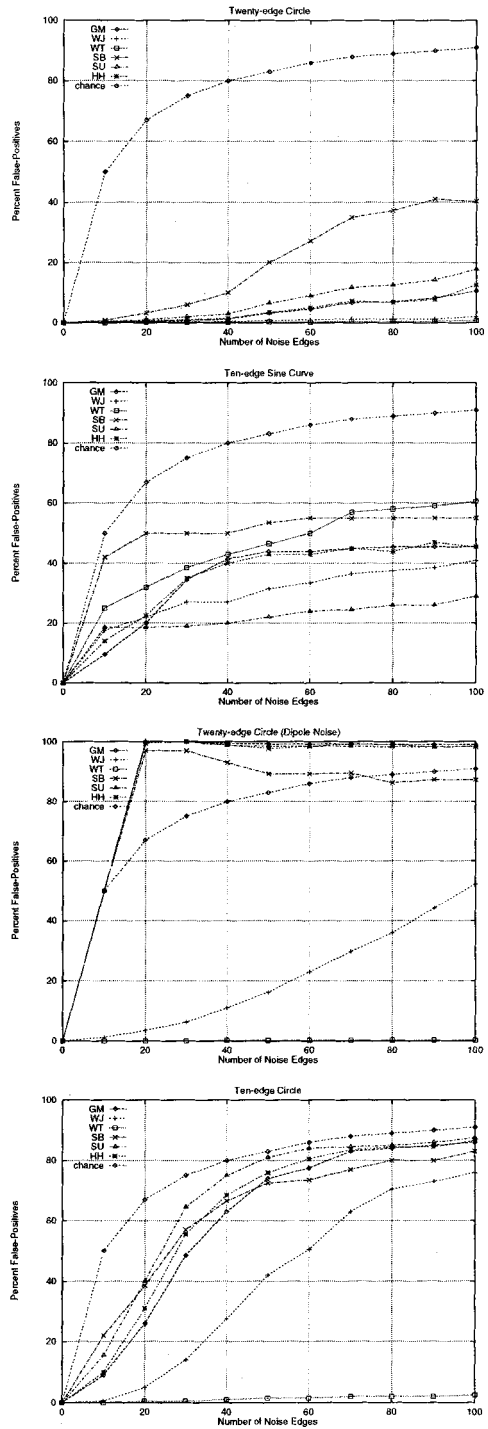
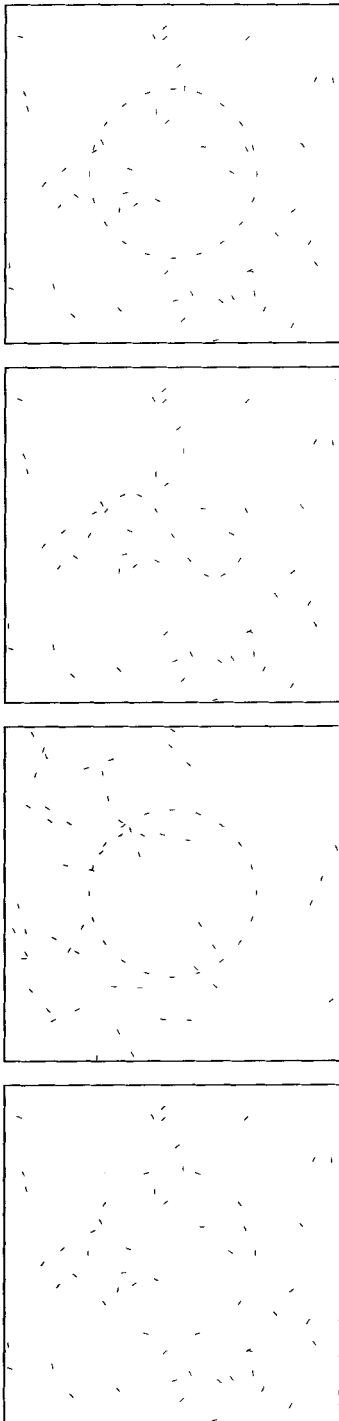
is useful to compare this to the saliency which Shashua and Ullman assigns to a curve, which is given by the following geometric series:

$$\Phi(\Gamma) = \sum_{j=1}^{\infty} \prod_{i=1}^j P'(\Gamma_{i+1} | \Gamma_i)$$

Shashua and Ullman desired a saliency measure which favored long, smooth contours yet converged to a finite value for contours of infinite length (i.e., for closed contours). Unfortunately, the rate at which this series converges depends critically on the values of the  $P'(\Gamma_{i+1} | \Gamma_i)$ . If the transition probabilities are too small, the series will converge too rapidly (and the measure becomes local and greedy). Conversely, if they are too large, the series will converge too slowly. In summary, we see that the geometric mean has the properties Shashua and Ullman wanted but lacks other undesirable properties.

## 4 Results

The first comparison used test patterns which consisted of short oriented edges spaced uniformly around the perimeter of a circle in a background of edges



**Fig. 5.** (a) twenty-edge circle. (b) ten-edge sine curve. (c) twenty-edge circle (dipole noise) (d) ten-edge circle. All patterns are shown at a noise-level of fifty.

with random positions and orientations (see Figure 5 (a)).<sup>8</sup> We computed the saliency of both shape and noise edges using each of the six measures: SU, HH, SB, GM, WJ, and WT. The edges were then sorted in ascending order based on their saliencies. The saliency of the most salient edge is  $\phi_1$  and the saliency of the least salient edge is  $\phi_n$ . Given  $m$  shape edges, we define a false-positive as a noise edge which is assigned a saliency larger than  $\phi_{m+1}$ . The false-positive rate for each measure was computed for patterns consisting of different numbers of shape and noise edges. The false-positive rate for each combination (e.g., 20 shape edges and 70 noise edges) was estimated by averaging the false-positive rate for ten trials using different noise patterns.<sup>9</sup> The right half of Figure 5(a) is a plot of the percentage false-positives versus the number of noise edges for the twenty-edge circle.

All of the measures perform reasonably well (less than 10% false-positive rate) at the low noise-levels (40 noise edges or less). At higher noise-levels, the performance of the measures begins to diverge. It is interesting that GM significantly outperforms SB, since GM is essentially one iteration of SB. We speculate that the false-positive rate is increased by additional relaxation-labeling steps using the non-directional operator. We also observe that GM performs comparably to HH—even though the HH measure is significantly more expensive to compute. Finally, at the lower signal-to-noise ratios, WJ and WT have significantly lower false-positive rates.

The second comparison was identical to the first except that the shape edges formed an open-ended sine curve (see Figure 5 (b)). The right half of Figure 5 (b) is a plot of the percentage false-positives versus the number of noise edges for the ten-edge sine curve. The relatively poor performance of the WT measure compared to the other measures can be attributed to its explicit reliance on closure. Nevertheless, it still outperforms the SB measure for higher signal-to-noise ratios and has an error rate comparable to that of SB (i.e., within 5%) at lower signal-to-noise ratios. As in the previous comparison, the performance of GM and HH are nearly identical. The false-positive rates of these measures is somewhat larger than that of the WJ measure. The SU measure had the best performance.

In the third comparison, we used a background consisting of correlated (i.e., dipole) noise (see Figure 5(c)). A dipole consists of two collinear edges separated by a gap of size equal to the distance between successive edges of the circle. Because the two edges forming a dipole are collinear, the affinity between the edges forming a dipole is greater than between adjacent circle edges. Consequently, it is impossible to distinguish noise edges from shape edges using purely

---

<sup>8</sup> The radius of the circle equals 32 and the noise edges are uniformly distributed within a square of size 64.

<sup>9</sup> We wanted to ensure that HH was not unfairly penalized because of the inherent difficulty of solving the combinatorial optimization problem. We therefore computed the value of  $\mathbf{y}^T \mathbf{A} \mathbf{y}$  for the perfect segmentation and accepted a trial only when the simulated annealing procedure returned a greater or equal value. After ten failed attempts, we restarted that trial with a new noise-pattern.

local measures. Indeed, all of the measures but WJ and WT have nearly a 100% false-positive rate. In the case of the SU measure, this is because (for gaps of this size) the geometric series is dominated by the first two terms.<sup>10</sup>

In the fourth comparison (see Figure 5(d)), we used a ten-edge circle. This is a challenging pattern because the sampling rate is so low—only one edge per 36 degrees of circumference. Most of the measures perform poorly, even at relatively high signal-to-noise ratios. For a noise-level of 80, the GM, SU and HH measures are performing almost at chance, or 90% false-positive rate. The SB and WJ measures perform slightly better, with false-positive rates of 80% and 70%, respectively. In contrast, the false-positive rate for WT is under 5%.

Our intention in the last comparison was to test the saliency measures on a collection of “real images” but to do so in a way which would allow meaningful error rates to be estimated. In the past, when new grouping methods have been proposed, their performance has not been systematically compared to others from the literature. Although the proposed methods are typically demonstrated on two or three “real images,” because the computational goal is often not well defined, performance is impossible to gauge. Consequently, it is unclear whether or not the methods represent genuine improvements in the state of the art.

We decided to construct test patterns from pairs of real images in such a way that performance on the saliency problem could be objectively measured. Nine different fruits and vegetables were placed in front of a uniformly colored background (three of these are shown in Figure 6(a-c)). This allowed their silhouettes to be extracted using straightforward methods. The orientation at points uniformly spaced along the silhouette was then estimated using a robust line fitting technique (see Figure 6(d-f)).

Next, we selected nine images of natural texture from the MIT Media Lab texture database (these of these are shown in Figure 6(g-i)). The Canny edge detector was applied to a  $64 \times 64$  block from each texture and the resulting edges were filtered on contrast to create a set of nine masking patterns consisting of approximately 800 edges each (see Figure 6(j-l)).

Edges from the nine fruit and vegetable silhouettes (signal) and nine natural texture masking patterns (noise) were then combined to construct a set of 405 test patterns. These patterns represent all 81 silhouette and texture combinations at five different signal-to-noise ratios (see Figure 6(m-o)).<sup>11</sup> Each of the six saliency measures was run on all of the test patterns and false-positive rates were computed as before. The results are plotted in Figure 7. For a signal-to-noise ratio of 0.2, the false-positive rate for the SB measure is 72% (i.e., 8% better than chance performance). The false-positive rates for SU, GM, HH and WJ are all approximately 50%. In contrast, the false-positive rate for the WT measure is 20% (i.e., 60% better than chance performance). Furthermore, after the signal-to-noise ratio is reduced by a factor of two, the false-positive rate for the WT measure remains under 50%.

<sup>10</sup> For the thirty-edge circle, the geometric series converges more slowly. Presumably, SU would continue to improve (relative to the other measures) as the size of the gaps decreases.

<sup>11</sup> The texture edges are undersampled to achieve a given signal-to-noise ratio.

## 5 Conclusion

In this paper, we introduced a new measure of perceptual saliency and quantitatively compared its ability to detect natural shapes in cluttered backgrounds to five previously proposed measures. The saliency measure is based on the distribution of closed random walks through the edges. We computed false-positive rates in classifying edges as signal or noise for a large set of test figures. In almost every case, the new measure significantly outperforms previous measures.

## 6 Appendix

In this Appendix, we give the analytic expression for the affinity function used in the comparisons (see [13] for its derivation).<sup>12</sup> We define the affinity,  $P_{ji}$ , between two directed edges,  $i$  and  $j$ , to be:

$$P_{ji} \equiv P'(j|i) = \int_0^\infty dt P(j|i;t) \approx F P(j|i;t_{opt})$$

where  $P(j|i;t)$  is the probability that a particle which begins its stochastic motion at  $(x_i, y_i, \theta_i)$  at time 0 will be at  $(x_j, y_j, \theta_j)$  at time  $t$ . The affinity between two edges is the value of this expression integrated over stochastic motions of all durations,  $P'(j|i)$ . This integral is approximated analytically using the method of steepest descent. The approximation is the product of  $P$  evaluated at the time at which the integral is maximized (i.e.,  $t_{opt}$ ), and an extra factor,  $F$ . The expression for  $P$  at time  $t$  is:

$$P(j|i;t) = \frac{3 \exp[-\frac{6}{Tt^3}(at^2 - bt + c)] \cdot \exp(-\frac{t}{\tau})}{\sqrt{\pi^3 T^3 t^\tau / 2}}$$

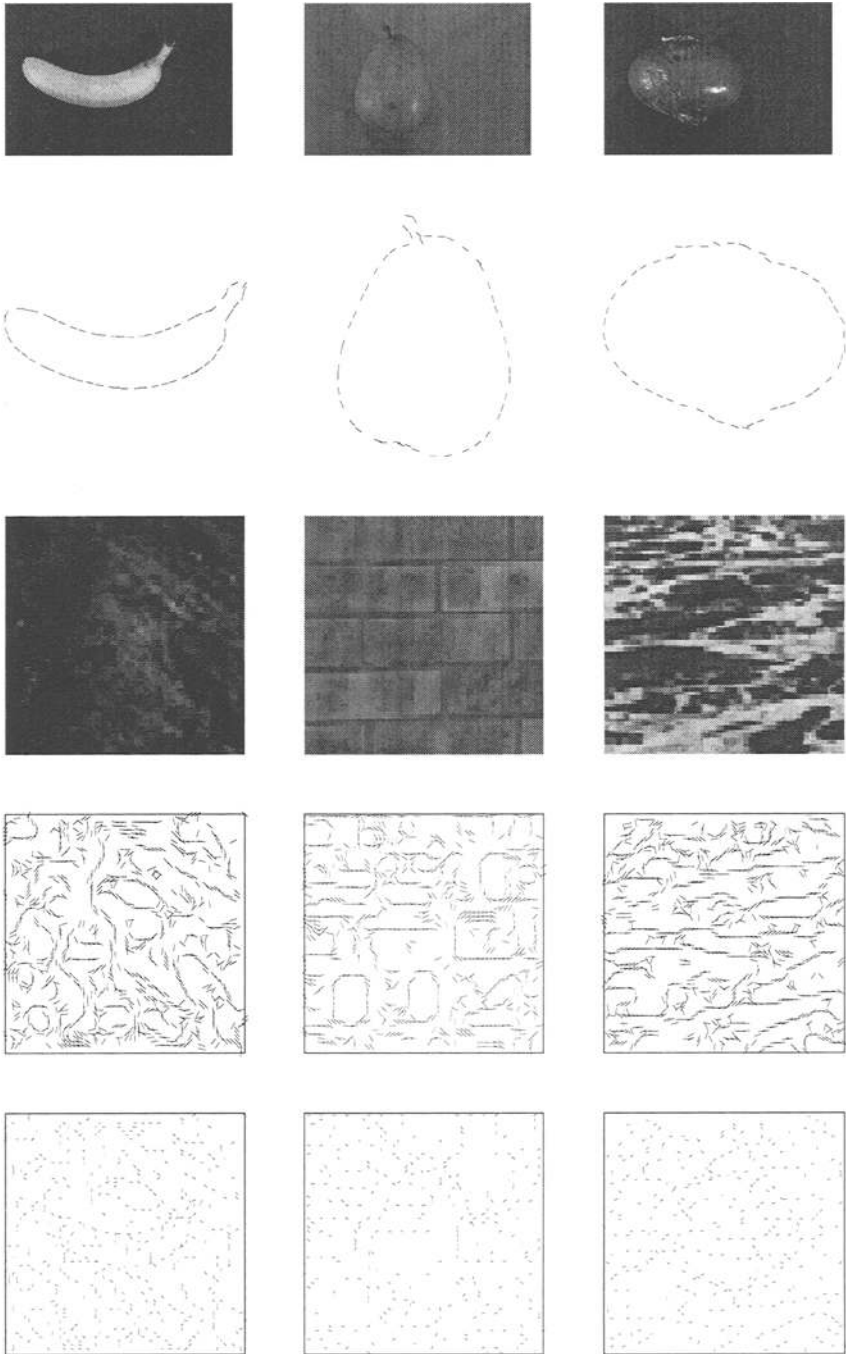
where

$$\begin{aligned} a &= [2 + \cos(\theta_j - \theta_i)] / 3 \\ b &= [x_{ji}(\cos \theta_j + \cos \theta_i) + y_{ji}(\sin \theta_j + \sin \theta_i)] / v \\ c &= (x_{ji}^2 + y_{ji}^2) / v^2 \end{aligned}$$

for  $x_{ji} = x_j - x_i$  and  $y_{ji} = y_j - y_i$ . The parameters  $T$ ,  $\tau$  and  $v$  determine the distribution of shapes (where  $T$  is the diffusion coefficient,  $\tau$  is particle half-life and  $v$  is speed). In all of our experiments,  $T = 0.002$ ,  $\tau = 5.0$  and  $v = 1$ . The expression for  $P$  should be evaluated at  $t = t_{opt}$ , where  $t_{opt}$  is real, positive, and satisfies the following cubic equation:

$$-7t^3/4 + 3(at^2 - 2bt + 3c)/T = 0$$

<sup>12</sup> For a derivation of a related affinity function, see the recent paper of Sharon, Brandt and Basri[12].



**Fig. 6.** (a-c) Banana, pear, red onion (d-f) Banana edges, pear edges, red onion edges (g-i) Terrain, brick, water (j-l) Terrain edges, brick edges, water edges (m-o) Banana with terrain mask, pear with brick mask, red onion with water mask.

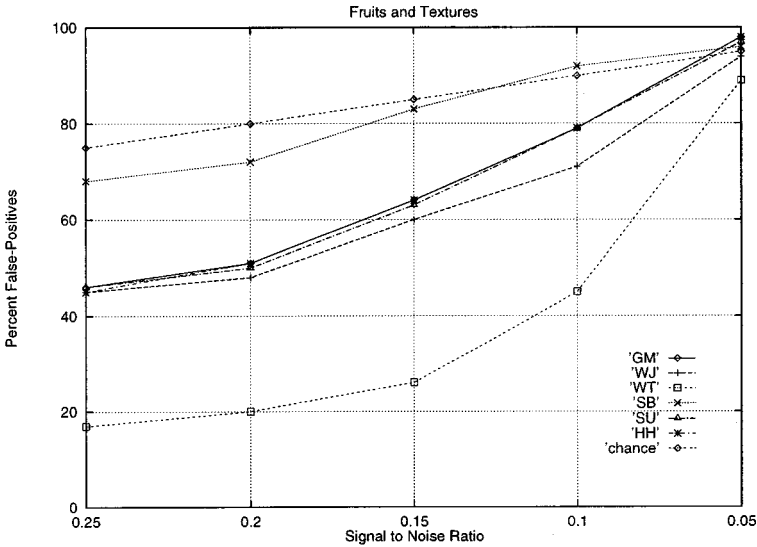


Fig. 7. False-positive rate for fruit and vegetable silhouettes with natural texture backgrounds.

If more than one real, positive root exists, then the root maximizing  $P(j|i;t)$  is chosen.<sup>13</sup> Finally, the extra factor  $F$  is:

$$F = \sqrt{2\pi t_{opt}^5 / [12(3c - 2bt_{opt})/T + 7t_{opt}^3/2]}$$

For our purposes here, we ignore the  $\exp(-t/\tau)$  factor in the steepest descent approximation for  $t_{opt}$ . We note that by increasing  $v$ , the distribution of contours can be uniformly scaled.

**Acknowledgments** The authors wish to thank Hong Pan for providing the silhouettes of fruits and vegetables. Thanks also to Shyjan Mohamud and Majd Sakr for their assistance in running experiments and plotting results. Finally, we are grateful to David Jacobs and Michael Langer for many helpful discussions.

## References

1. Alter, T. and R. Basri, Extracting Salient Contours from Images: An Analysis of the Saliency Network, *Proc. IEEE Conf. on Comp. Vision and Pattern Recognition (CVPR '96)*, pp. 13-20, San Francisco, CA, 1996.

<sup>13</sup> For a discussion on solving cubic equations, see [8].



2. Grossberg, S., and E. Mingolla, Neural Dynamics of Form Perception: Boundary Completion, Illusory Figures, and Neon Color Spreading, *Psychological Review* **92**, pp. 173-211, 1985.
3. Guy, G. and G. Medioni, Inferring Global Perceptual Contours from Local Features, *Intl. Journal of Computer Vision* **20**, pp. 113-133, 1996.
4. Horn, B.K.P., The Curve of Least Energy, MIT AI Lab Memo No. 612, MIT, Cambridge, Mass., 1981.
5. Héroult, L. and R. Horaud, Figure-Ground Discrimination: A Combinatorial Optimization Approach, *IEEE Trans. on Pattern Analysis and Machine Intelligence* **15**, pp. 899-914, 1993.
6. Montanari, U., On the Optimal Detection of Curves in Noisy Pictures, *Comm. of the Assoc. for Computing Machinery* **14**, pp. 335-345, 1971.
7. Mumford, D., Elastica and Computer Vision, *Algebraic Geometry and Its Applications*, Chandrajit Bajaj (ed.), Springer-Verlag, New York, 1994.
8. Press, W.H., Flannery, B.P., Teukolsky, S.A., and W.T. Vetterling, *Numerical Recipes in C*, Cambridge University Press, 1988.
9. Rosenfeld, A., Hummel R., and S. Zucker, Scene Labeling by Relaxation Operations, *IEEE Trans. on Systems Man and Cybernetics* **6**, pp. 420-433, 1976.
10. Sarkar, S. and K. Boyer, Quantitative Measures of Change based on Feature Organization: Eigenvalues and Eigenvectors, *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR '96)*, pp. 478-483, San Francisco, CA, 1996.
11. Shashua, A. and S. Ullman, Structural Saliency: The Detection of Globally Salient Structures Using a Locally Connected Network, *2nd Intl. Conf. on Computer Vision*, Clearwater, FL, 1988.
12. Sharon, E., Brandt, A., and R. Basri, Completion Energies and Scale, *Proc IEEE Conf. Computer Vision and Pattern Recognition (CVPR '97)*, pp. 884-890, San Juan, Puerto Rico, 1997.
13. Thornber, K.K. and L.R. Williams, Analytic Solution of Stochastic Completion Fields, *Biological Cybernetics* **75**, pp. 141-151, 1996.
14. Ullman, S., Filling-in the Gaps: The Shape of Subjective Contours and a Model for Their Generation, *Biological Cybernetics* **21**, pp. 1-6, 1976.
15. Williams, L.R., Wang, T. and K.K. Thornber, Computing Stochastic Completion Fields in Linear-Time Using a Resolution Pyramid, *Proc. of 7th Intl. Conf. on Computer Analysis of Images and Patterns (CAIP '97)*, Kiel, Germany, 1997.
16. Williams, L.R. and D.W. Jacobs, Stochastic Completion Fields: A Neural Model of Illusory Contour Shape and Saliency, *Neural Computation* **9**, pp. 849-870, 1997.
17. Yen, S. and L. Finkel, "Pop-Out" of Salient Contours in a Network Based on Striate Cortical Connectivity, *Investigative Ophthalmology and Visual Science (ARVO)*, Vol. **37**, No. 3, 1996.