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

LANCE R. WILLIAMS*
*Dept. of Computer Science, University of New Mexico, Albuquerque, NM 87131, USA*

KARVEL K. THORNBER
*NEC Research Institute, 4 Independence Way, Princeton, NJ 08540, USA*

**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.

**Keywords:** contours, saliency, closure, grouping

## 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 Shashua and Ullman, 1988) 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. 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 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 cocircularity (Guy and Medioni, 1996; Hérault and Horaud, 1993; Parent and Zucker, 1989; Ullman, 1976). 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 (Horn, 1981; Shashua and Ullman, 1988). The affinity between two directed edges is inversely related to the energy, $\int_{\Gamma} ds \, (\alpha \kappa^2(s) + \beta)$, 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 (Mumford, 1994; Thornber and Williams, 1996; Williams and Jacobs, 1997). 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., $P'(j \mid i)$. 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 (Mumford, 1994; Williams and Jacobs, 1997). It follows that the function $P'(j \mid i)$ 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 similiar 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. Every edge has a single orientation
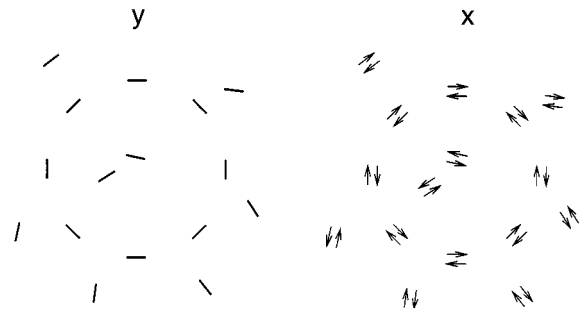


*Figure 1.* Every edge has a single *orientation* and two distinct *directions*. Its orientation, $\theta$, is an angular quantity in the range, $0$ to $\pi$. Its directions are $\theta$ and $\theta + \pi$. Left: A vector, $\mathbf{y}$, with $n$ components, can be used to associate a single value, $y_i$, with each of $n$ edges. Right: A vector, $\mathbf{x}$, of length $2n$, can be used to associate two values, $x_i$ and $x_{\bar{\imath}}$, with each of $n$ edges. These values correspond to the two distinct directions.
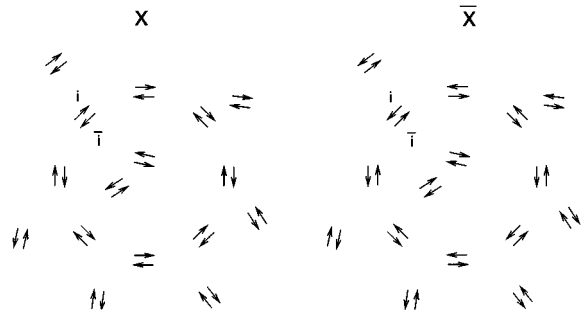


*Figure 2.* The vectors, $\mathbf{x}$ and $\bar{\mathbf{x}}$, are identical except for a permutation which exchanges the values associated with opposite directions. That is, $x_i = \bar{x}_{\bar{\imath}}$ and $x_{\bar{\imath}} = \bar{x}_i$.

and two distinct directions. Its orientation, $\theta$, is an angular quantity in the range, $0$ to $\pi$. Its directions are $\theta$ and $\theta + \pi$. 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 vector, $\mathbf{x}$, has $2n$ components while the vector, $\mathbf{y}$, has $n$ components (see Fig. 1). The vectors $\mathbf{x}$ and $\bar{\mathbf{x}}$ are identical except for a permutation which exchanges the values associated with opposite directions, i.e., $x_i = \bar{x}_{\bar{\imath}}$ and $x_{\bar{\imath}} = \bar{x}_i$. For example, if $x_i$ is a value associated with an edge of direction, $\theta$, then $\bar{x}_i$ is a value associated with the same edge, but in the opposite direction, $\theta + \pi$ (see Fig. 2).

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 $\mathbf{A}$ to represent the affinities between all pairs of oriented edges and the
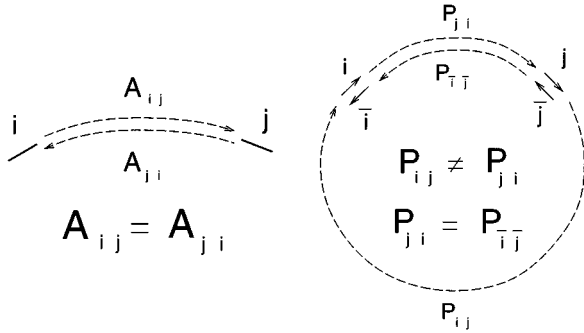
*Figure 3.*    Left: The $n \times n$ matrix **A** can be used to represent the affinity values between all pairs of oriented edges. Right: The $2n \times 2n$ matrix, **P**, is needed to represent the affinity values between all pairs of directed edges. While the affinity oriented edge $i$ has for $j$ 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., $\mathbf{A} = \mathbf{A^T}$) but in general, $P_{ij} \neq P_{ji}$. Although not symmetric, **P** exhibits another kind of symmetry. If we use the subscript $\bar{i}$ to denote the opposite direction to $i$ then $P_{ij} = P_{\bar{j}\bar{i}}$ and $P_{ji} = P_{\bar{i}\bar{j}}$. This is termed *time-reversal* symmetry.

$2n \times 2n$ matrix **P** to represent the affinity values between all pairs of directed edges (see Fig. 3). 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., $\mathbf{A} = \mathbf{A^T}$) but in general, $P_{ij} \neq P_{ji}$.

Although not symmetric, **P** exhibits another kind of symmetry. If we use the subscript $\bar{i}$ to denote the opposite direction to $i$ then $P_{ij} = P_{\bar{j}\bar{i}}$ and $P_{ji} = P_{\bar{i}\bar{j}}$. This is termed *time-reversal* symmetry. For the purposes of our comparison, we will define $A_{ij}$ to be $\max(P_{ij}, P_{\bar{i}j}, P_{i\bar{j}}, P_{\bar{i}\bar{j}})$, 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 (1988) 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 (1971), they described a saliency measure which could be computed by a local parallel network.[1] 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 (1996) have done an extensive analysis of Shashua and Ullman's method and give expressions for the saliency measure for the case of continuous curves.[2] The saliency of a directed edge $i$ equals the maximum of the saliencies of all continuous curves, $\Gamma$, which begin at that edge:

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

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 $\sigma(.)$ 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 $\kappa^2(.)$ 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.[3] 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.

### 2.2.  *Hérault and Horaud (HH)*

Hérault and Horaud (1993) 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, $-E_{\text{saliency}} - E_{\text{constraint}}$:

$$\min_{\mathbf{y}} \left( -\frac{1}{2} \mathbf{y}^{\mathbf{T}} \mathbf{H} \mathbf{y} - \mathbf{b}^{\mathbf{T}} \mathbf{y} \right) \quad \text{for } \mathbf{y} \in \{-1, +1\}^n$$

where $H_{ij} = A_{ij} - \alpha$ and $b_i = \sum_j (A_{ij} - \alpha)$. The affinity function used by Hérault and Horaud is based on co-circularity, smoothness and proximity. Hérault 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 Hérault-Horaud with measures lacking a comparable parameter difficult. Therefore (for the comparison) we decided to maximize $E_{\text{saliency}}$ over 0–1 solution vectors with exactly $m$ components equal to 1:

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

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 Hérault-Horaud measure.

### 2.3. Sarkar and Boyer (SB)

Sarkar and Boyer (1996) 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 and perpendicularity. The affinity function we used in the comparison is the same one we used with the other methods (i.e., equivalent to only a subset of the relations proposed by Sarkar and Boyer). Given an affinity matrix, $\mathbf{A}$, Sarkar and Boyer propose that the saliency vector, $\mathbf{y}$, maximizes

the Raleigh Quotient:

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

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

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

This measure can also be optimized using the following recurrence equation:

$$y_i^{(t+1)} = \frac{\sum_j A_{ij} y_j^{(t)}}{\sum_j \sum_k A_{jk} y_k^{(t)}}$$

which has been independently proposed as a saliency computation by Yen and Finkel (1996) and by Perona and Freeman (1998). From linear algebra, we know that the vector $\mathbf{y}$ will converge to the eigenvector associated with the largest positive real eigenvalue of $\mathbf{A}$. Viewed this way, we see that $\mathbf{A}$ is being used as a linear relaxation labeling operator and that maximizing the Raleigh Quotient is equivalent to solving a linear relaxation labeling problem as defined by Rosenfeld, Hummel and Zucker (1976).

### 2.4. Guy and Medioni (GM)

Guy and Medioni (1996) 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 $\mathbf{A}$ and a constant input vector.

## 2.5.  *Williams and Jacobs (WJ)*

Williams and Jacobs (1997) 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 the (component-wise) product of $\mathbf{x}$ and $\bar{\mathbf{x}}$ 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.[4]

## 3.    A New Measure (WT)

We define the salience of an edge to be the relative number of closed random walks which visit that edge. By random walk, we mean a sequence of edges visited by a particle with edge-to-edge transition probabilities given by the matrix, $\mathbf{P}$. By closed random walk, we mean a random walk which begins and ends at the same edge. Although it is important to distinguish between (1) random walks; and (2) the paths in the plane followed by a particle when $\theta(t)$ is a Brownian motion, these two notions are intimately related. This is because the probability that a particle located at edge $i$ at time-step $t$ will be at edge $j$ at time-step $t + 1$ is defined to be the sum over the probabilities of all paths between $i$ and $j$, i.e., $P_{ji} \equiv P'(j \mid i)$. 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, $\mathbf{P}$. 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:

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

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.[5] 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:

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

While there are other expressions for $\lambda$, the above expression is significant because it makes explicit the relationship between error in $\mathbf{x}$ and error in $\lambda$.[6] However, unlike the Raleigh Quotient for symmetric matrices, $\mathbf{x}$ does not maximize this expression. Furthermore, although $\mathbf{x}$ is a fixed-point, there is no guarantee that a process which starts at a random vector and repeatedly applies the recurrence equation will converge to $\mathbf{x}$.[7]

Recall that our stated goal was to compute the relative number of closed random walks through every edge. We defined a closed random walk to be a random walk which begins and ends at the same edge. Unfortunately, because $\mathbf{P}$ is not a Markov matrix, the eigenvector with largest positive real eigenvalue does not represent a distribution of closed random walks—only a small subset of the random walks contributing to $\mathbf{x}$ are actually closed. The problem which concerns us is how to characterize this subset.

To begin, we observe that any random walk of infinite length beginning and ending at edge $i$ can be exhaustively decomposed into an infinite number of closed random walks each of which visits edge $i$ exactly once. Because (after cyclic permutation) any visit to edge $i$ is also the beginning and ending state of an infinite length random walk, we conclude that the relative number of closed random walks which visit edge

$i$ and the relative number of random walks of infinite length which begin and end at edge $i$ are in one-to-one correspondence. The following expression gives the relative number of random walks of infinite length which begin and end at edge $i$:

$$c_i = \lim_{t \to \infty} \frac{\mathbf{P}_{ii}^t}{\sum_j \mathbf{P}_{jj}^t}$$

To evaluate the above expression, we first divide $\mathbf{P}$ by its largest real positive eigenvalue, $\lambda$, and distribute the limit over the numerator and denominator, which yields

$$c_i = \frac{\lim_{s \to \infty} \left(\frac{\mathbf{P}}{\lambda}\right)_{ii}^s}{\lim_{t \to \infty} \sum_j \left(\frac{\mathbf{P}}{\lambda}\right)_{jj}^t}$$

To evaluate the numerator and denominator, we observe that $\mathbf{P}$ can be written as $\mathbf{X D \bar{X}^T}$ where the columns of $\mathbf{X}$ are $\mathbf{x}_\mu / (\bar{\mathbf{x}}_\mu^T \mathbf{x}_\mu)^{\frac{1}{2}}$, the right eigenvectors of $\mathbf{P}$ and $\mathbf{D}$ is diagonal with the corresponding eigenvalues, $\mu$, on the diagonal. From time-reversal symmetry, $\lambda \mathbf{x} = \mathbf{P x}$ implies $\lambda \bar{\mathbf{x}} = \mathbf{P^T \bar{x}}$. Consequently, the columns of $\bar{\mathbf{X}}$ are $\bar{\mathbf{x}}_\mu / (\bar{\mathbf{x}}_\mu^T \mathbf{x}_\mu)^{\frac{1}{2}}$, the left eigenvectors of $\mathbf{P}$. Hence, for $t$ sufficiently large,

$$\lim_{t \to \infty} \left(\frac{\mathbf{P}}{\lambda}\right)^t = \lim_{t \to \infty} \mathbf{X} \left(\frac{\mathbf{D}}{\lambda}\right)^t \bar{\mathbf{X}}^T$$
$$= \lim_{t \to \infty} \sum_\mu \frac{\mathbf{x}_\mu \left(\frac{\mu}{\lambda}\right)^t \bar{\mathbf{x}}_\mu^T}{\bar{\mathbf{x}}_\mu^T \mathbf{x}_\mu}$$
$$= \frac{\mathbf{x}_\lambda \bar{\mathbf{x}}_\lambda^T}{\bar{\mathbf{x}}_\lambda^T \mathbf{x}_\lambda}$$

This is because $\lambda$ is the largest real positive eigenvalue of $\mathbf{P}$ and therefore $\lambda > |\mu|$ for all eigenvalues, $\mu$, except $\mu = \lambda$. Applying this result to the denominator of the expression for $c_i$ yields

$$\lim_{t \to \infty} \sum_j \left(\frac{\mathbf{P}}{\lambda}\right)_{jj}^t = \sum_j \frac{(\mathbf{x} \bar{\mathbf{x}}^T)_{jj}}{\bar{\mathbf{x}}^T \mathbf{x}} = 1$$

Consequently, the relative number of closed random walks through edge $i$ is given by the numerator

$$c_i = \lim_{s \to \infty} \left(\frac{\mathbf{P}}{\lambda}\right)_{ii}^s = \frac{(\mathbf{x} \bar{\mathbf{x}}^T)_{ii}}{\bar{\mathbf{x}}^T \mathbf{x}} = \frac{x_i \bar{x}_i}{\sum_j x_j \bar{x}_j}$$

To within a constant factor, this is simply $x_i \bar{x}_i$. The whole distribution is given by the (component-wise)

product of $\mathbf{x}$ and $\bar{\mathbf{x}}$. In effect, at each edge we are constructing the cartesian product of the set of random walks bridging edges in all past-times and the set consisting of their time-reversed counterparts.[8] This forms a set of closed random walks.

## 4.  Directionality and Tangent Continuity

In the Yen and Finkel (1996) saliency computation, the support for oriented edge $i$ due to all other oriented edges $j$ is given by the linear relaxation labeling update step, $y_i^{(t+1)} = \sum_j A_{ji} y_j^{(t)} / \sum_j \sum_k A_{jk} y_k^{(t)}$. Because Yen and Finkel's intention was to model the visual cortex, the $\mathbf{y}^{(t)}$ vector represents a fixed lattice of discrete positions and orientations. The update step can be viewed as convolution with a large kernel filter followed by normalization.[9] The symmetry of the $\mathbf{A}$ matrix manifests itself in the plane as mirror image symmetry in the kernel (see Fig. 4). The state of the computation at any given time is summarized by $\mathbf{y}^{(t)}$, which represents a distribution of random walks of length $t$. In general, this iteration will converge to the eigenvector, $\mathbf{y}$, associated with the largest positive real eigenvalue of $\mathbf{A}$, i.e., it computes the SB measure.

For illustrative purposes, it is worth considering a network analogous to that described by Yen and Finkel, but based on a lattice of discrete positions and directions (i.e., $\mathbf{x}^{(t)}$ instead of $\mathbf{y}^{(t)}$) and using a
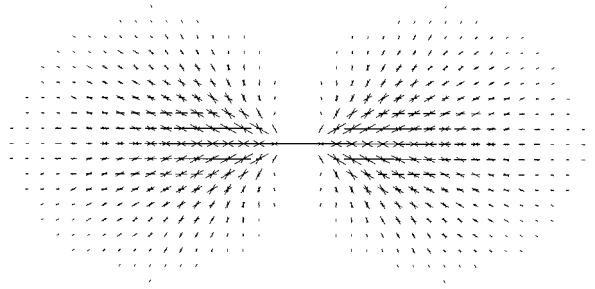


*Figure 4.* In the Yen and Finkel (1996) saliency computation, the support for oriented edge $i$ due to all other oriented edges $j$ is given by the linear relaxation labeling update step, $y_i^{(t+1)} = \sum_j A_{ji} y_j^{(t)} / \sum_j \sum_k A_{jk} y_k^{(t)}$. Because Yen and Finkel's intention was to model the visual cortex, the $\mathbf{y}^{(t)}$ vector represents a fixed lattice of discrete positions and orientations. The update step can be viewed as convolution with a large kernel filter followed by normalization. The symmetry of the $\mathbf{A}$ matrix manifests itself in the plane as mirror image symmetry in the kernel. In general, this iteration will converge to the eigenvector, $\mathbf{y}$, associated with the largest positive real eigenvalue of $\mathbf{A}$, i.e., it computes the SB measure.
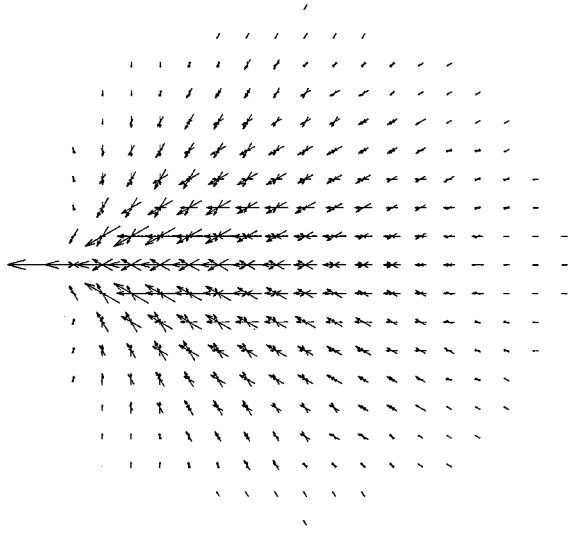
*Figure 5.* Hypothetical saliency network is based on a lattice of discrete positions and directions (i.e., $\mathbf{x}^{(t)}$ instead of $\mathbf{y}^{(t)}$) and uses a non-symmetric linear relaxation labeling operator (i.e., $\mathbf{P}$ instead of $\mathbf{A}$). In this network, the support for directed edge $i$ due to all other directed edges $j$ is given by the linear relaxation labeling update step, $x_i^{(t+1)} = \sum_j P_{ji} x_j^{(t)} / \sum_j \sum_k P_{jk} x_k^{(t)}$. Because $\mathbf{P}$ is not symmetric, this iteration is not guaranteed to converge. The salience 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 $\mathbf{P}$. This quantity represents the relative number of closed random walks through edge $i$.

non-symmetric linear relaxation labeling operator (i.e., $\mathbf{P}$ instead of $\mathbf{A}$), see Fig. 5. In this network, the support for directed edge $i$ due to all other directed edges $j$ is given by the linear relaxation labeling update step, $x_i^{(t+1)} = \sum_j P_{ji} x_j^{(t)} / \sum_j \sum_k P_{jk} x_k^{(t)}$. As before, the state of the computation at any given time is summarized by $\mathbf{x}^{(t)}$, which represents a distribution of random walks of length $t$. However, because $\mathbf{P}$ is not symmetric, there is no guarantee that this process will converge to $\mathbf{x}$, the eigenvector with largest positive real eigenvalue. Furthermore, although $\mathbf{x}$ is a fixed-point, $\mathbf{x}$ alone does not represent a distribution of closed random walks. The salience of edge $i$ is given by the product of the $i$th components of the right and left eigenvectors, $x_i$ and $\bar{x}_i$. It is the product which represents the relative number of closed random walks through edge $i$.

We observe that the use of directions and a non-symmetric linear relaxation operator is essential—even for visual patterns consisting of non-directional elements, e.g., Gabor patches of even phase. While the intermediate states of both networks at time $t$ (i.e., $\mathbf{y}^{(t)}$ and

$\mathbf{x}^{(t)}$) can be interpreted as distributions of random walks of length $t$, the random walks of the Yen and Finkel network contain *cusps* (i.e., reversals in direction). The fundamental problem is that without representing directions in the state vector, the linear relaxation labeling process cannot remember the direction of travel of the random walk in the previous time-step—it only knows its orientation. Because the linear relaxation labeling operator is symmetric, at the next time-step, the random walk is equally likely to continue in either direction. One of these directions preserves tangent continuity. The other introduces a cusp. This is illustrated in Fig. 6. After the initial iteration, this process tends to increase the salience of the noise edges as much as the signal edges. It explains why the performance of Guy and Medioni's saliency computation (essentially one iteration of linear relaxation labeling using the operator $\mathbf{A}$ and a constant input vector) is superior to that of Sarkar and Boyer. It also explains why WT outperforms both.

The importance of directionality in enforcing tangent continuity in computations involving contours seems not to be generally recognized in the human vision community. See for example, the paper of Field et al. (1993), which suggests that the phenomenon of "pop-out" of salient contours is evidence for an
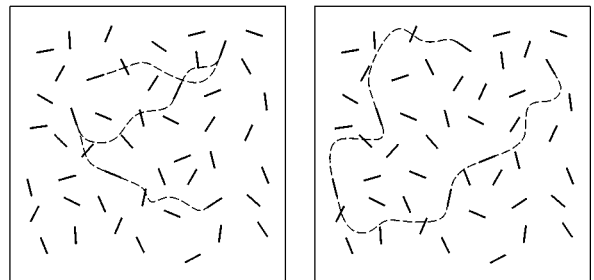


*Figure 6.* The explicit representation of two directions (i.e., $\mathbf{x}$ and $\bar{\mathbf{x}}$) and the use of a *non-symmetric* linear relaxation labeling operator (i.e., $\mathbf{P}$) is essential if the intermediate states of the relaxation labeling process (i.e., $\mathbf{x}^{(t)}$, $\mathbf{x}^{(t+1)}$, 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., $\mathbf{A}$) 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 $\mathbf{A}$ and a constant input vector) is superior to that of Sarkar and Boyer. It also explains why WT outperforms both.

"association field" which looks very much like Fig. 4. A notable exception is the "bipole" cell of the boundary contour system of Grossberg and Mingolla (1985), which (like our model) combines inputs from opposite directions.

## 5.    Contour Saliency

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:

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

satisfies $\det(P - \lambda I) = 0$. This is the *geometric mean* of the transition probabilities in the closed path.[10] 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 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.

## 6.    Results

### 6.1.    Saliency Maps for Simple Patterns

In order to gain some insight into the strengths and weaknesses of the various measures, it will be useful to apply them to a simple test pattern consisting of edges from a circle (either thirty, twenty, or ten uniformly spaced samples) in a background of one hundred edges of random position and orientation.[11] The three test patterns are shown in Fig. 7(a–c).

We will first look at the performance of the Shashua and Ullman (SU) measure. In order to visualize the saliencies assigned to each edge, the edges are displayed as rectangles with lengths and widths proportional to the raw saliency values. The raw saliency map for the thirty edge circle is shown in Fig. 8(a). It is clear that the SU measure assigns significantly higher saliencies to the edges of the circle than to the background edges. This would allow the circle to be segmented from its background simply by choosing an appropriate saliency threshold. The raw saliency map for the twenty edge circle is shown in Fig. 8(b). In this case, the saliencies of the circle edges are comparable to those of the background, making segmentation of the circle by thresholding impossible. Note also that one of the two most salient edges belongs to the background. This is an edge, which, simply through chance, lies almost exactly on the circle with the correct
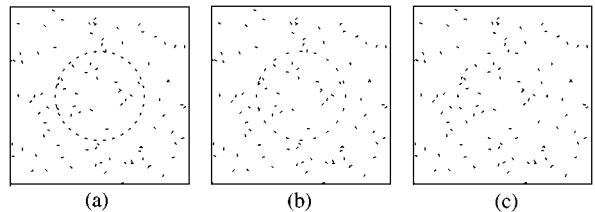


(a)                    (b)                    (c)

*Figure 7.*    (a) Thirty edge circle in a background of one hundred noise edges; (b) twenty edge circle and (c) ten edge circle.


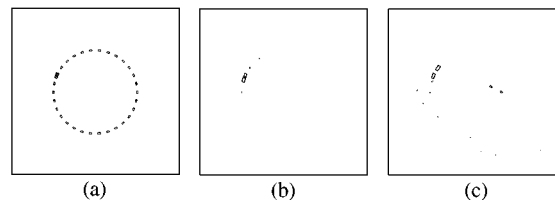
(a)                    (b)                    (c)

*Figure 8.*    (a) Raw saliencies computed by SU for thirty edge circle; (b) raw saliencies computed by SU for twenty edge circle and (c) raw saliencies computed by SU for ten edge circle.
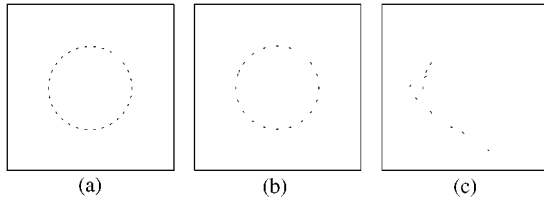
*Figure 9.*  (a) Figure edges computed by HH for thirty edge circle; (b) figure edges computed by HH for twenty edge circle and (c) figure edges computed by HH for ten edge circle.

orientation (approximately at the 10 o'clock position). It is unreasonable to expect any measure to be able to distinguish this edge from the edges of the circle. Finally, the performance of the SU measure on the ten edge circle (see Fig. 8(c)) is also poor.

We next applied the segmentation method of Hérault and Horaud (HH) to the same three test patterns. It is important to note that we are solving the modified optimization problem described previously. That is, we are optimizing $\mathbf{y^T A y}$ over all vectors, $\mathbf{y} \in \{0, 1\}^n$ and $|\mathbf{y}| = m$ where $n$ is the total number of edges and $m$ is the number of figure edges. Because it is given the number of figure edges, HH possesses a considerable advantage over the other measures and these results should be interpreted accordingly. The segmentation for the thirty and twenty edge circles are shown in Fig. 9(a) and (b). With the exception of omitting one circle edge at 1 o'clock to permit inclusion of the spurious 10 o'clock edge, HH computes a perfect segmentation. However, the results on the ten edge circle (see Fig. 9(c)) show that the method has failed to segment the circle from its background.

In order to better visualize the large range of saliency values computed by the method of Sarkar and Boyer (SB), the lengths and widths of the rectangles are drawn proportional to $\log(1.0 + 10^6 \cdot x_i)$, where $x_i$ is the saliency of edge $i$. Figure 10(a) shows the log saliencies for the thirty edge circle computed using the SB method. In general, the edges of the circle are assigned

greater saliencies than the edges of the background. However, we observe that the saliency values on the upper left portion of the circle are significantly larger than those on the lower right. If the eigenvector with largest positive real eigenvalue is interpreted as a limiting distribution of random walks between the edges, we see that this distribution is dominated by random walks (with reversals in direction) through the spurious edge at the 10 o'clock position. Because the SB measure does not enforce tangent continuity or closure, the effect of a single unfortuitously placed edge can be profound. In Fig. 10(b) and (c) the asymmetry becomes very pronounced as the sampling of the circle becomes less frequent. The consequence is that the SB measure has failed to isolate the circle from its background, even in a case where a very simple method like WJ has little problem (see Fig. 13(b)).[12]

Figure 11(a–c) shows the raw saliency values for the three circles computed using the method of Guy and Medioni (GM). Like the SU measure, the GM measure assigns significantly higher saliency values to the edges of the thirty edge circle than to the background edges. However, also like the SU measure, the GM meaure performs more poorly on the twenty and ten edge circles. Overall, the similiarities between the saliencies computed by the SU measure and those computed by the GM measure are quite striking. We speculate that, for contours defined by relatively few edges of negligible length separated by large gaps, the largest $P_{ij}$ are relatively small (e.g., approximately 0.1). Consequently, the geometric series computed by SU is dominated by its first term, which equals $\max_j P_{ij}$. In turn, because the spatial attenuation of the affinity function is rapid, $\sum_j P_{ij} \approx \max_j P_{ij}$, i.e., the sum of the affinities from all edges is dominated by the maximum affinity from all edges.

Given the similiarity of the saliencies computed by the SU and GM measures on the thirty, twenty, and ten edge circles, it is useful to include an example where the saliency values computed using the two methods are
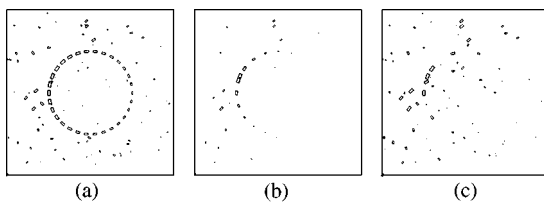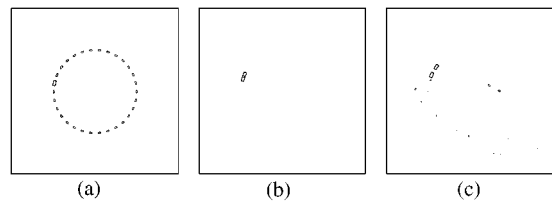


*Figure 10.*  (a) log Saliencies computed by SB for thirty edge circle; (b) log saliencies computed by SB for twenty edge circle and (c) log saliencies computed by SB for ten edge circle.



*Figure 11.*  (a) Raw saliencies computed by GM for thirty edge circle; (b) raw saliencies computed by GM for twenty edge circle and (c) raw saliencies computed by GM for ten edge circle.
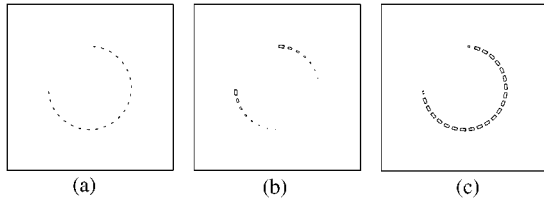
*Figure 12.* (a) An incomplete circle with twenty three edges; (b) raw saliency values computed using the SU measure and (c) raw saliency values computed using the GM measure.

very different. Figure 12(a) shows an incomplete circle with twenty three edges. There are no noise edges. Figure 12(b) shows the raw saliency values computed using the SU measure. Edges near the two ends of the incomplete circle are assigned the highest saliencies, and edges far from the ends are assigned the smallest saliencies. In this figure, the magnitude of $P_{ij}$ for adjacent $i$ and $j$ is greater than one. Consequently, $\sum_{n=1}^{22} P_{ij}^n \gg \sum_{n=1}^{11} P_{ij}^n$. This explains the difference in relative saliency. Figure 12(c) shows the raw saliency values computed using the GM measure. Except for the edges on the ends, which are assigned somewhat lower values, most edges are assigned comparable saliencies.

Figure 13(a) and (b) show the log saliency maps for the thirty and twenty edge circles computed using the method of Williams and Jacobs (WJ). With the exception of the spurious 10 o'clock edge, all of the background edges are assigned saliencies of negligible magnitude. The contrast between the saliencies computed by WJ and those computed by GM is quite striking. The difference in discrimination power is due solely to the use of directionality (i.e., **P** and **x** vs. **A** and **y**) and the (component-wise) multiplication of the vector **x** by the vector **x̄** in WJ. This multiplication enforces the constraint that an edge must form a bridge between two others. The result is saliencies with a greater range of magnitudes than possessed by the components of **x** or **x̄** alone. However, this constraint is not enough
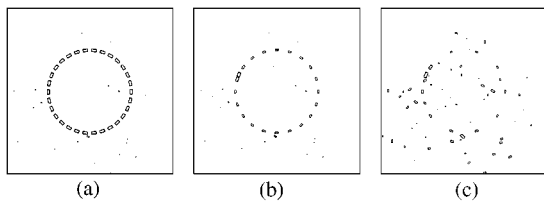


*Figure 13.* (a) Log saliencies computed by WJ for thirty edge circle; (b) log saliencies computed by WJ for twenty edge circle and (c) log saliencies computed by WJ for ten edge circle.
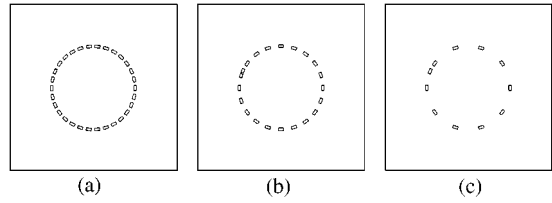


*Figure 14.* (a) Log saliencies computed by WT for thirty edge circle; (b) log saliencies computed by WT for twenty edge circle and (c) log saliencies computed by WT for ten edge circle.

to discriminate edges of the ten edge circle from those of the background. The log saliency map computed by the WJ measure for the ten edge circle is shown in Fig. 13(c). A significant number of background edges have saliencies comparable to those assigned to edges of the circle. Consequently, segmentation of the circle from its background using thresholding is not possible.

Finally, Fig. 14(a–c) show the log saliency maps for the three circles computed using the WT measure. With the exception of the spurious 10 o'clock edge, all of the background edges are assigned saliencies of negligible magnitude. This holds even for the ten edge circle—which no other measure was able to segment from the background.

### 6.2. False Positive Rates for Simple Patterns

The first quantitative comparison used test patterns which consisted of short oriented edges spaced uniformly around the perimeter of a circle in a background of edges with random positions and orientations (see Fig. 15(a)). 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 salience of the most salient edge is $\phi_1$ and the salience 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 salience 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.[13] The right half of Fig. 15(a) is a plot of the percentage false-positives versus the number of noise edges for the twenty-edge circle.

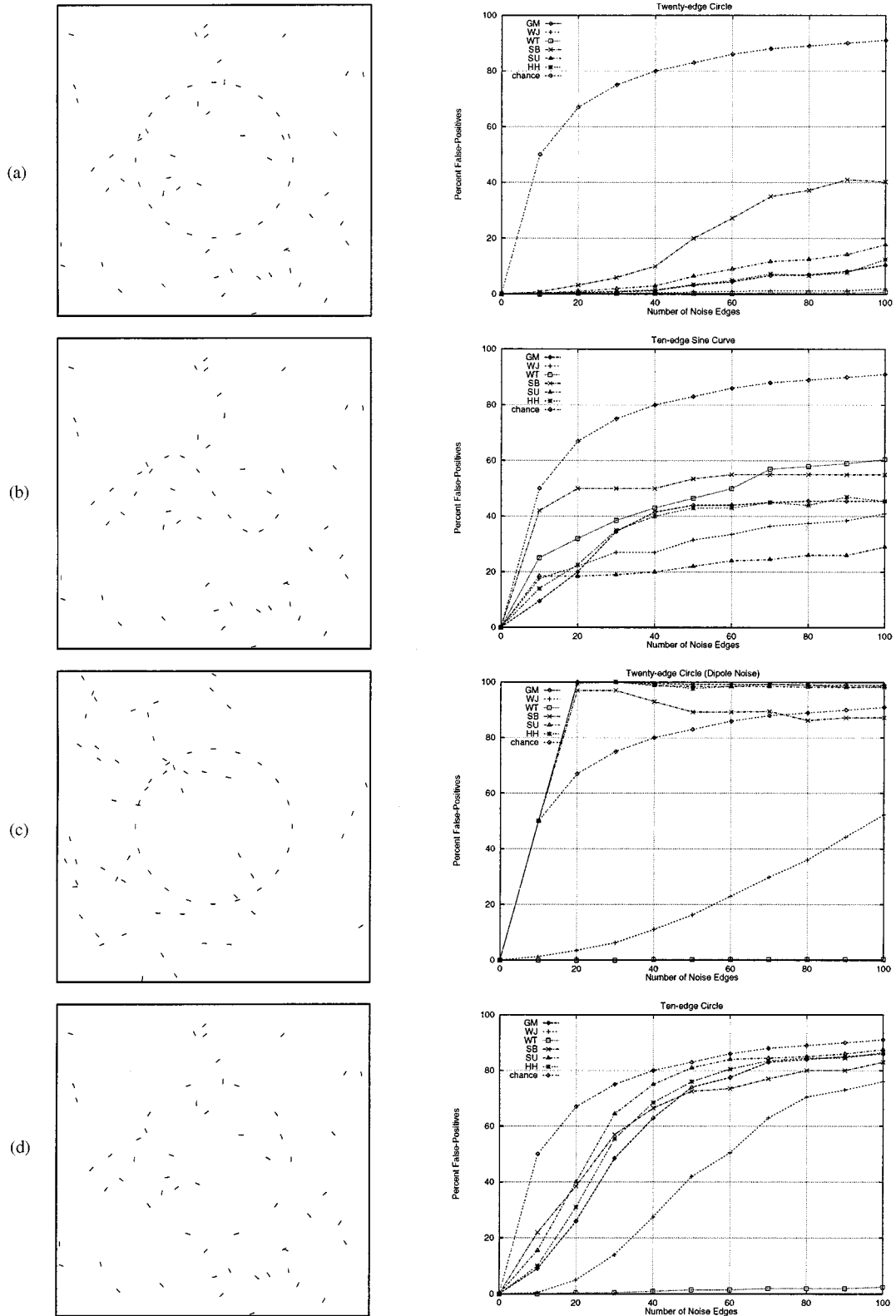All of the measures perform reasonably well (less than 10% false-positive rate) at the low noise-levels (40

*Figure 15.*    (a) Twenty edge circle; (b) ten edge sine curve; (c) twenty edge circle (dipole noise) and (d) ten edge circle. All patterns are shown at a noise-level of fifty.

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 Fig. 15(b)). The right half of Fig. 15(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 Fig. 15(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 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.[14]

In the fourth comparison (see Fig. 15(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%.

### 6.3. Fruit and Texture Patterns

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 Fig. 16(a–c)). This allowed their silhouettes
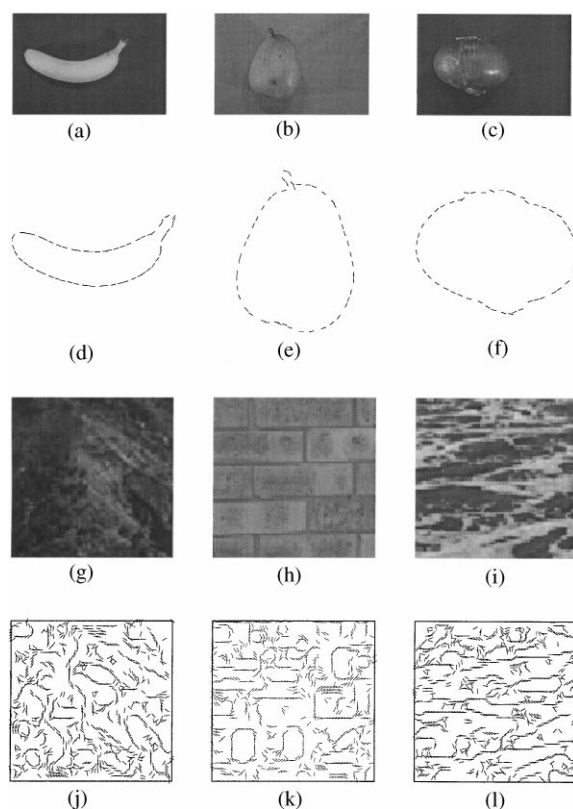


*Figure 16.* (a–c) Banana, pear, red onion, (d–f), banana edges, pear edges, red onion edges, (g–i) terrain, brick, water and (j–l) terrain edges, brick edges, water edges.

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 Fig. 16(d–f)). Because the fruits and vegetables varied widely in size, the maximum of the $x$ and $y$ dimensions of the silhouette was determined (i.e., the bounding square) and the set of edges was rescaled to an absolute size of $32 \times 32$.

To serve as background, we selected nine images of natural texture from the MIT Media Lab texture database (three of these are shown in Fig. 16(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 Fig. 16 (j–l)). Test patterns were constructed by insetting the fruit and vegetable silhouettes into the center $32 \times 32$ regions of the size $64 \times 64$ natural textures. Figure 17(a) displays the banana with terrain background, Fig. 17(b) displays the pear with brick background, and Fig. 17(c) displays the red onion with water background. Because the actual fruits and vegetables varied widely in size, the distances between adjacent edges of their silhouettes after rescaling is also somewhat variable. For example, because the banana was somewhat larger than the pear, the edges of its silhouette are more closely spaced after rescaling. This explains why some methods perform consistently better on one test pattern than on another.

The raw saliency maps computed by the SU measure for the three fruit and texture test patterns are shown in Fig. 18(a–c). The figure edges computed by HH are shown in Fig. 19(a–c). The log saliency maps computed by the SB measure for the three fruit and texture test patterns are shown in Fig. 20(a–c). The raw saliency maps computed by the GM measure are shown in Fig. 21(a–c). The log saliency maps computed by the WJ measure are shown in Fig. 22(a–c). Finally, the log saliency maps computed by the WT measure are
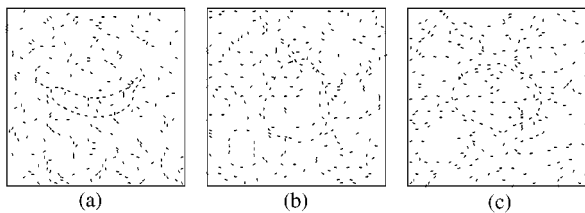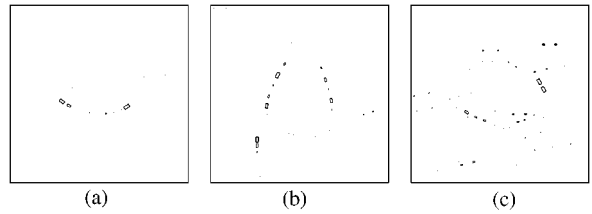


*Figure 18.* (a) Raw saliencies computed by SU for banana with terrain mask; (b) raw saliencies computed by SU for pear with brick mask and (c) raw saliencies computed by SU for red onion with water mask.
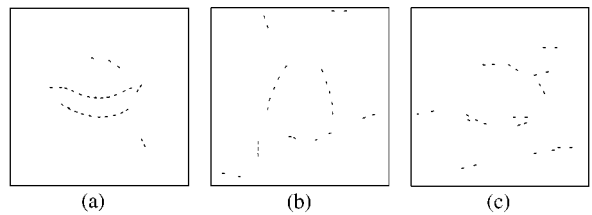


*Figure 19.* (a) Figure edges computed by HH for banana with terrain mask; (b) figure edges computed by HH for pear with brick mask and (c) figure edges computed by HH for red onion with water mask.
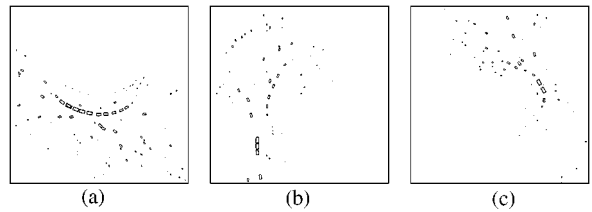


*Figure 20.* (a) Log saliencies computed by SB for banana with terrain mask; (b) log saliencies computed by SB for pear with brick mask and (c) log saliencies computed by SB for red onion with water mask.
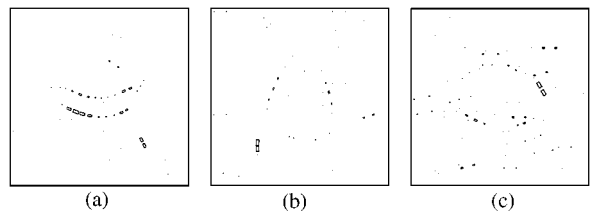


*Figure 21.* (a) Raw Saliencies computed by GM for banana with terrain mask; (b) raw saliencies computed by GM for pear with brick mask and (c) raw saliencies computed by GM for red onion with water mask.



*Figure 17.* (a) Banana with terrain mask; (b) pear with brick mask and (c) red onion with water mask.

shown in Fig. 23(a–c). We observe that, qualitatively, the performance of the various measures on the fruit and texture test patterns is similiar to that observed for the circle test patterns.
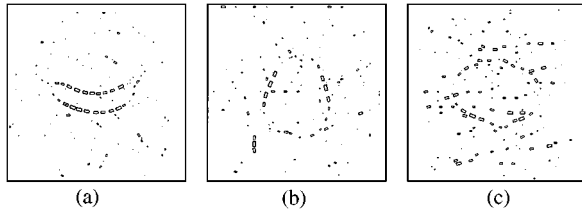
*Figure 22.* (a) Log saliencies computed by WJ for banana with terrain mask; (b) log saliencies computed by WJ for pear with brick mask and (c) log saliencies computed by WJ for red onion with water mask.
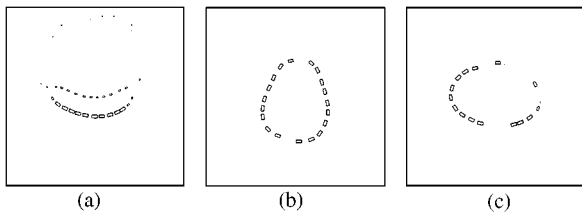


*Figure 23.* (a) Log saliencies computed by WT for banana with terrain mask; (b) log saliencies computed by WT for pear with brick mask and (c) log saliencies computed by WT for red onion with water mask.

For a quantitative comparison, edges from the nine fruit and vegetable silhouettes (signal) and nine natural texture masking patterns (noise) were combined to construct a set of 405 test patterns. The texture edges are undersampled to achieve a given signal-to-noise ratio. These patterns represent all 81 silhouette and texture combinations at five different signal-to-noise ratios (see Fig. 16(m–o)).[15] 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 Fig. 24. For a signal-to-noise ratio of 0.2, the false-
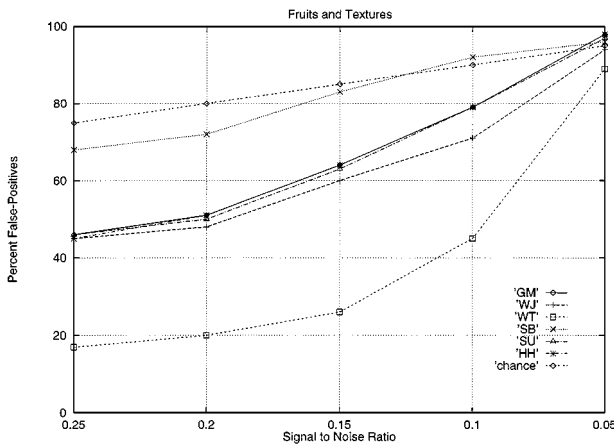


*Figure 24.* False-positive rate for fruit and vegetable silhouettes with natural texture backgrounds.

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%.

## 7. 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. We observe that the explicit representation of two edge directions and a non-symmetric affinity matrix are necessary to enforce tangent continuity in salient contours. Apart from our method (WT), only the method of Shashua and Ullman (SU) possesses this property. However, unlike our method, the SU method does not explicitly enforce contour closure. Since smooth, closed contours are judged to be more salient by human observers (Field et al., 1993; Kovacs and Julesz, 1993), we conjecture that these two reasons underly the improved performance of our method relative to previous methods.

## Appendix

In this Appendix, we give the analytic expression for the affinity function used in the comparisons (see (Thornber and Williams, 1996) for its derivation).[16] We define the affinity, $P_{ji}$, between two directed edges, $i$ and $j$, to be:

$$P_{ji} \equiv P'(j \mid i) = \int_0^\infty dt\, P(j \mid i; t) \approx FP(j \mid i; t_{\text{opt}})$$

where $P(j \mid i; t)$ 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, $P'(j \mid i)$. 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., $t_{\mathrm{opt}}$), and an extra factor, $F$. The expression for $P$ at time $t$ is:

$$P(j \mid i; t) = \frac{3 \exp\left[ -\frac{6}{Tt^3}(at^2 - bt + c)\right] \cdot \exp\left(\frac{-t}{\tau}\right)}{\sqrt{\pi^3 T^3 \, t^7/2}}$$

where

$$a = [2 + \cos(\theta_j - \theta_i)] / 3$$
$$b = [x_{ji}(\cos\theta_j + \cos\theta_i) + y_{ji}(\sin\theta_j + \sin\theta_i)]/\gamma$$
$$c = \left(x_{ji}^2 + y_{ji}^2\right)/\gamma^2$$

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

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

If more than one real, positive root exists, then the root maximizing $P(j \mid i; t)$ is chosen.[17] Finally, the extra factor $F$ is:

$$F = \sqrt{\frac{2\pi t_{\mathrm{opt}}^5}{12(3\,c - b\,t_{\mathrm{opt}})/T + 7t_{\mathrm{opt}}^3/2}}$$

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

## Acknowledgments

## Notes

1. 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 $i$.

2. Anyone who is interested in understanding the Shashua and Ullman saliency computation in greater detail is encouraged to read Alter and Basri's very helpful paper.

3. A typical value for $\exp(-\int_{s_1}^{s_i} dt\, (\kappa^2(t) - \ln\rho))$ is 0.1, so that $\Phi(\Gamma) = 1 + 0.1 + 0.01 + \cdots = 1.\overline{11}$. In such cases, ninety-nine percent of the saliency is contributed by the first two terms.

4. It is also worth noting that the WJ measure can be computed very efficiently using a multi-resolution method. See (Williams et al., 1997)

5. 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-probability matrix must sum to one. Consequently, the largest eigenvalue also equals one.

6. Specifically, it shows that if $\mathbf{x}$ were in error by $\delta\mathbf{x}$, the calculated $\lambda$ would be in error by only $\delta\mathbf{x}$ squared.

7. For this reason, in all of our experiments, we compute the eigenvector with largest positive real eigenvalue using EISPACK.

8. Those random walks which visit edges of opposite direction and in reverse order.

9. Probably the best example of this style of visual computation is the Parent and Zucker (1989) network, which uses (non-linear) relaxation labeling to improve local measurements of contour tangent and curvature. The relaxation labeling operator represents constraints between discrete tangent and curvature labels at all locations in the plane. Although the Parent and Zucker network and the Yen and Finkel network are similiar at an algorithmic level, they compute very different functions. Indeed, the goal of the former is more accurately described as "sharpening" than "saliency."

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

11. The radius of the circle equals 16 and the noise edges are uniformly distributed within a square of size 64.

12. It is worth pointing out (again) that Sarkar and Boyer did not design their method for problems of this sort. Their intention was to segment aerial photos of construction sites. It is likely that tangent continuity and closure are less important for that domain. However, Yen and Finkel (1996) and Perona and Freeman (1998) describe the same measure and demonstrate it using simple test patterns very much like the ones shown here.

13. 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 $\mathbf{y}^{\mathbf{T}}\mathbf{A}\mathbf{y}$ 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.

14. 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.

15. The original fruit and texture images and files containing the silhouette and texture edges can be downloaded from http://www.cs.unm.edu/~williams/saliency.html.

16. For a derivation of a related affinity function, see the recent paper of Sharon, Brandt and Basri (1997).

17. For a discussion on solving cubic equations, see Press et al., 1988.

In general, this quantity will not converge to a finite value as $t$ goes to infinity.

## References

Alter, T. and Basri, R. 1996. Extracting salient contours from images: An analysis of the saliency network. In *Proc. IEEE Conf. on Comp. Vision and Pattern Recognition (CVPR '96)*, San Francisco, CA, pp. 13–20.

Field, D.J., Hayes, A., and Hess, R.F. 1993. Contour integration by the human visual system: Evidence for a local association field. *Vision Research*, 33:173–193.

Grossberg, S. and Mingolla, E. 1985. Neural dynamics of form perception: Boundary completion, illusory figures, and neon color spreading. *Psychological Review*, 92:173–211.

Guy, G. and Medioni, G. 1996. Inferring global perceptual contours from local features. *Intl. Journal of Computer Vision*, 20:113–133.

Horn, B.K.P. 1981. The curve of least energy, MIT AI Lab Memo No. 612, MIT, Cambridge, MA.

Hérault, L. and Horaud, R. 1993. Figure-ground discrimination: A combinatorial optimization approach. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 15:899–914.

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*, 90:7495–7497.

Montanari, U. 1971. On the optimal detection of curves in noisy pictures. *Comm. of the Assoc. for Computing Machinery*, 14:335–345.

Mumford, D. 1994. Elastica and computer vision. In *Algebraic Geometry and Its Applications*, Chandrajit Bajaj (Ed.), Springer-Verlag: New York.

Parent, P. and Zucker, S.W. 1989. Trace inference, curvature consistency and curve detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11:823–889.

Perona, P. and Freeman, W. 1998. A factorization approach to grouping. In *Proc. of the 5th European Conf. on Computer Vision (ECCV '98)*, Freiburg, Germany.

Press, W.H., Flannery, B.P., Teukolsky, S.A., and Vetterling, W.T. 1988. *Numerical Recipes in C*, Cambridge University Press.

Rosenfeld, A., Hummel, R., and Zucker, S. 1976. Scene labeling by relaxation operations. *IEEE Trans. on Systems Man and Cybernetics* 6:420–433.

Sarkar, S. and Boyer, K. 1996. Quantitative measures of change based on feature organization: Eigenvalues and eigenvectors. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR '96)*, San Francisco, CA, pp. 478–483.

Shashua, A. and Ullman, S. 1988. Structural saliency: The detection of globally salient structures using a locally connected network. In *2nd Intl. Conf. on Computer Vision (ICCV '88)*, Clearwater, FL.

Sharon, E., Brandt, A., and Basri, R. 1997. Completion energies and scale. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR '97)*, San Juan, Puerto Rico, pp. 884–890.

Thornber, K.K. and Williams, L.R. 1996. Analytic solution of stochastic completion fields. *Biological Cybernetics*, 75:141–151.

Ullman, S. 1976. Filling-in the gaps: The shape of subjective contours and a model for their generation. *Biological Cybernetics*, 21:1–6.

Williams, L.R., Wang, T., and Thornber, K.K. 1997. Computing stochastic completion fields in linear-time using a resolution pyramid. In *Proc. of 7th Intl. Conf. on Computer Analysis of Images and Patterns (CAIP '97)*, Kiel, Germany.

Williams, L.R. and Jacobs, D.W. 1997. Stochastic completion fields: A neural model of illusory contour shape and salience. *Neural Computation*, 9:849–870.

Yen, S. and Finkel, L. 1996. "Pop-Out" of salient contours in a network based on striate cortical connectivity. *Investigative Opthalmology and Visual Science (ARVO)*, 37(3):5293.