ARTICLE Communicated by David Mumford

Orientation, Scale, and Discontinuity as Emergent Properties of Illusory Contour Shape

Lance R. Williams
Department of Computer Science, University of New Mexico, Albuquerque, NM 87131, U.S.A.

Karvel K. Thornber
NEC Research Institute, Princeton, NJ 08540, U.S.A.

A recent neural model of illusory contour formation is based on a distribution of natural shapes traced by particles moving with constant speed in directions given by Brownian motions. The input to that model consists of pairs of position and direction constraints, and the output consists of the distribution of contours joining all such pairs. In general, these contours will not be closed, and their distribution will not be scaleinvariant. In this article, we show how to compute a scale-invariant distribution of closed contours given position constraints alone and use this result to explain a well-known illusory contour effect.

1 Introduction

Mumford (1994) has proposed that the distribution of illusory contour shapes can be modeled by particles traveling with constant speed in directions given by Brownian motions. More recently, Williams and Jacobs (1997a, 1997b) introduced the notion of a stochastic completion field, the distribution of particle trajectories joining pairs of position and direction constraints, and showed how it could be computed in a local parallel network. They argued that the mode, magnitude, and variance of the completion field are related to the observed shape, salience, and sharpness of illusory contours.

Unfortunately, the Williams and Jacobs model, as described, has some shortcomings. Recent psychophysics suggests that contour salience is greatly enhanced by closure (Kovacs & Julesz, 1993). Yet, in general, the distribution computed by the Williams and Jacobs model does not consist of closed contours, nor is it scaleinvariant; doubling the distances between the constraints does not produce a comparable completion field of double the size without a corresponding doubling of particle speeds. However, the Williams and Jacobs model contains no intrinsic mechanism for speed selection. The speeds (like the directions) must be specified a priori.

In this article, we show how to compute a scale-invariant distribution of closed contours given position constraints alone. The significance of this is

twofold. First, it provides insight into how the output of neurons with purely isotropic receptive fields like those of lateral geniculate nucleus (LGN) might be combined to produce orientation-dependent responses like those exhibited by neurons in V1 and V2. Second, it suggests that the responses of these neurons are not just input for long-range contour grouping processes, but instead are emergent properties of such computations.

2 A Discrete-State, Continuous-Time Random Process

2.1 Shape Distribution. Mumford (1994) observed that the probability distribution of boundary completion shapes could be modeled by a Fokker-Planck equation of the following form:

$$\frac{\partial P}{\partial t} = -\gamma \cos\theta \frac{\partial P}{\partial x} - \gamma \sin\theta \frac{\partial P}{\partial y} + \frac{\sigma^2}{2}\frac{\partial^2 P}{\partial t^2} - \frac{1}{\tau}P$$

This partial differential equation can be viewed as a set of independent advection equations in $\vec{x} = [x, y]^{\mathrm{T}}$ (the first and second terms) coupled in the θ dimension by the diffusion equation (the third term). The advection equations translate probability mass in direction θ, with constant speed γ, while the diffusion term models the Brownian motion in direction, with diffusion parameter σ. The combined effect of these three terms is that particles tend to travel in straight lines, but over time they drift to the left or right by an amount proportional to σ^2. Finally, the effect of the fourth term is that particles decay over time, with a half-life given by the decay constant τ. This represents our prior expectation on the lengths of gaps: most are quite short.

The Green's function, $G_\gamma(\vec{x}, \theta\,;\, t_1 \mid \vec{u}, \phi\,;\, t_0)$, gives the probability that a particle (traveling with speed γ) observed at position $\vec{u}$, and direction ϕ, at time t_0, will later be observed at position $\vec{x}$, and direction θ, at time t_1. It is the solution, $P(\vec{x}, \theta\,;\, t_1)$, of the Fokker-Planck initial value problem with initial value, $P(\vec{x}, \theta\,;\, t_0) = \delta(\vec{x} - \vec{u})\delta(\theta - \phi)$. The symmetries of the Green's function are summarized by the following equation:

$$G_\gamma(\vec{x}, \theta\,;\, t_1 \mid \vec{u}, \phi\,;\, t_0) = G_1(R_\phi(\vec{x} - \vec{u})/\gamma, \theta - \phi\,;\, t_1 - t_0 \mid 0, 0\,;\, 0)$$

where the function $R_\phi(.)$ rotates its argument by ϕ about the origin, $[0, 0]^{\mathrm{T}}$. Two of these symmetries are especially relevant to this article. The first of these is scale invariance:

$$G_\gamma(\vec{x}, \theta\,;\, t_1 \mid \vec{u}, \phi\,;\, t_0) = G_1(\vec{x}/\gamma, \theta\,;\, t_1 \mid \vec{u}/\gamma, \phi\,;\, t_0).$$

Scale invariance requires that the probability of a particle following a path between two edges be invariant under a transformation that scales both the speed of the particle and the distance between the two edges by the same factor. The second symmetry that concerns us here is time-reversal symmetry:

$$G_\gamma(\vec{x}, \theta\,;\, t_1 \mid \vec{u}, \phi\,;\, t_0) = G_\gamma(\vec{u}, \phi + \pi;\, t_1 \mid \vec{x}, \theta + \pi\,;\, t_0).$$

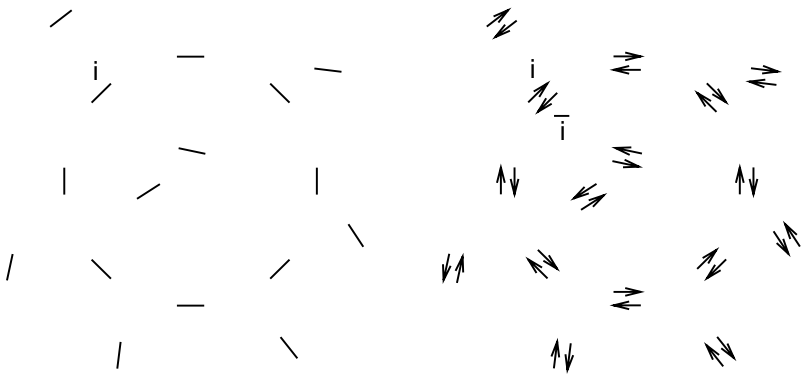

Figure 1: (Left) An input pattern consisting of N edges. Each edge i has a position $\vec{x}_i$ and an orientation θ_i. (Right) A set of $2N$ states, S. For each edge i in the input pattern, there are two states, i and $\bar{i}$, in S. The two states associated with an edge have the same position but are opposite in direction. That is, $\vec{x}_i = \vec{x}_{\bar{i}}$, but $\theta_i = \theta_{\bar{i}} + \pi$. These two states represent the two possible directions a particle undergoing a random motion can visit an edge of the input pattern.

Time-reversal symmetry requires that the probability of a particle following a path between two edges be invariant under a transformation that reverses the order and directions of the two edges.

2.2 Particles Visiting Edges. Consider an input pattern consisting of N edges. Each edge, i, has a position, $\vec{x}_i$, and an orientation, θ_i. Now construct a set of $2N$ states, S. For each edge i in the input pattern, there are two states, i and $\bar{i}$, in S. The two states associated with an edge have the same position but are opposite in direction; that is, $\vec{x}_i = \vec{x}_{\bar{i}}$ but $\theta_i = \theta_{\bar{i}} + \pi$. These two states represent the two possible directions a particle undergoing a random motion can visit an edge of the input pattern. Henceforward, we will refer to states simply as edges. (See Figure 1.)

Assuming that the shape of contours in the world can be modeled as the trajectories of particles undergoing random motions, we seek answers to questions like these:

- What is the probability that a contour begins at edge i and ends at edge j?
- What is the probability that a contour begins at edge i and reaches edge j and contains $n - 1$ other edges from the set S?

To begin, we use the Green's function to define an expression for the conditional probability,

$$P(j \mid i;\ t_1 - t_0) = G_1(\vec{x}_j, \theta_j;\ t_1 \mid \vec{x}_i, \theta_i;\ t_0),$$

that a particle (traveling with unit speed) observed at edge i at some unspecified time will subsequently be observed at edge j after time t_1 has elapsed. Integrating over all possible visitation times for edge j yields:

$$P(j \mid i) = \int_0^\infty dt_1 P(j \mid i\,;\; t_1).$$

This expression gives the probability that a particle observed at edge i will later be observed at edge j. Because there are no intermediate edges, we term this a path of length $n = 1$ joining i and j. In contrast, a path of length $n = 2$ joining i and j is a path that begins at edge i, next visits some unspecified intermediate edge k, and then ends at edge j. In addition to integrating over all possible visitation times, t_1, for edge j, the expression for $P^{(2)}(j \mid i)$ requires summing the probabilities of paths through all possible intermediate edges, k_1, and integrating over all possible visitation times, t_1, for edge k_1:

$$P^{(2)}(j \mid i) = \int_0^\infty dt_2 \int_0^{t_2} dt_1 \sum_{k_1} P(j \mid k_1\,;\; t_2 - t_1) P(k_1 \mid i\,;\; t_1),$$

where the limits of integration reflect the fact that t_2 is constrained to be greater than t_1. By extending the pattern, we can define a path of length n joining i and j to be a path that begins at edge i, next visits $n-1$ unspecified edges, and then ends at edge j. The expression for $P^{(n)}$ requires n sums over the set of possible intermediate edges and $n+1$ integrals over visitation times:

$$\begin{aligned} &P^{(n)}(j \mid i) \\ &\quad = \int_0^\infty dt_n \cdots \int_0^{t_2} dt_1 \sum_{k_n} \cdots \sum_{k_1} P(j \mid k_n\,;\; t_n - t_{n-1}) \cdots P(k_1 \mid i\,;\; t_1), \end{aligned}$$

where the limits of integration reflect the fact that t_m is constrained to be greater than t_{m-1} for $1 \leq m \leq n$. By reversing the order in which the integrals are evaluated, so that the first integral evaluated is over t_n and the last evaluated is over t_1, and making the necessary changes in the limits of integration, we get the following expression:

$$\begin{aligned} &P^{(n)}(j \mid i) \\ &\quad = \int_0^\infty dt_1 \cdots \int_{t_{n-1}}^\infty dt_n \sum_{k_n} \cdots \sum_{k_1} P(j \mid k_n\,;\; t_n - t_{n-1}) \cdots P(k_1 \mid i\,;\; t_1). \end{aligned}$$

We now substitute τ_m for $t_m - t_{m-1}$, for $1 \leq m \leq n$, and move the integrals inside the expression, so that each integral is immediately to the left of the conditional probability involving its variable of integration:

$$P^{(n)}(j \mid i) = \int_0^\infty d\tau_n \cdots \int_0^\infty d\tau_1 \sum_{k_n} \cdots \sum_{k_1} P(i \mid k_n\,;\; \tau_n) \cdots P(k_1 \mid i\,;\; \tau_1)$$

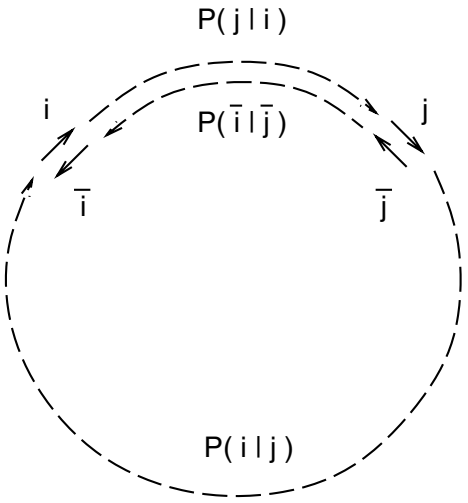

Figure 2: The matrix **P** represents the transition probabilities between pairs of edges. In general, $P(i \mid j) \neq P(j \mid i)$; that is, **P** is not symmetric. However, **P** possesses another form of symmetry: $P(i \mid j) = P(\bar{j} \mid \bar{i})$ and $P(j \mid i) = P(\bar{i} \mid \bar{j})$. This is termed time-reversal symmetry.

$$= \sum_{k_n} \cdots \sum_{k_1} \int_0^\infty d\tau_n P(j \mid k_n\,;\ \tau_n) \cdots \int_0^\infty d\tau_1 P(k_1 \mid i\,;\ \tau_1)$$
$$= \sum_{k_n} \cdots \sum_{k_1} P(j \mid k_n) \cdots P(k_1 \mid i).$$

The result is an expression for $P^{(n)}(j \mid i)$, the probability of a length n path joining i and j, purely in terms of probabilities of length one paths. Let us now define a $2N \times 2N$ matrix, **P**, where $P_{ji} = P(j \mid i)$, that is, P_{ji} is the probability of a length one path between edges i and j. Because $P(j \mid i) \neq P(i \mid j)$, the matrix, **P**, is not symmetric. However, by time-reversal symmetry, $P(j \mid i) = P(\bar{i} \mid \bar{j})$ where $\theta_{\bar{i}} = \theta_i + \pi$ and $\theta_{\bar{j}} = \theta_j + \pi$. (See Figure 2.) From the above expression, it follows that:

$$P^{(n)}(j \mid i) = \left(\mathbf{P}^n\right)_{ji}.$$

This result is significant because it shows that the probability of a particle with motion governed by a continuous time random process visiting n edges at n real valued times can be computed simply by taking the nth power of a matrix. The implication is that our analysis from this point on need not involve the continuous time random process; we need not write expressions involving integrals over continuous times. Rather, our analysis can be based entirely on the discrete-time random process with transition probabilities

specified by the matrix, **P**. Furthermore, any expressions we derive based on the discrete-time process will also apply to the underlying continuous-time process.

3 A Discrete-State, Discrete-Time Random Process

3.1 Edge Saliency. In section 2, we derived an expression for the probability that a particle moving with constant speed in a direction given by a Brownian motion will travel from edge i to edge j and visit $n-1$ intermediate edges. Significantly, this expression involved only the probabilities of length one paths:

$$P(j \mid i) = \int_0^\infty P(j \mid i\,;\; t)dt.$$

In effect, we reduced the problem of computing visitation probabilities for a continuous-time random process to the more tractable problem of computing visitation probabilities for a discrete-time random process. In the discrete-time process, the probability of a path of length n between edges i and j can be computed using the following recurrence equation:

$$P^{(n)}(j \mid i) = \sum_k P(j \mid k)P^{(n-1)}(k \mid i).$$

This equation allows us to define an expression for the relative number of contours that visit edge i and eventually return to edge i:

$$c_i = \lim_{n\to\infty} \frac{P^{(n)}(i \mid i)}{\sum_j P^{(n)}(j \mid j)}.$$

This quantity, which we term the edge saliency, is the relative number of closed contours through edge i. To evaluate the above expression, we first divide **P** by its largest real positive eigenvalue, λ, allowing us to distribute the limit over the numerator and denominator. This yields

$$c_i = \frac{\lim_{n\to\infty} \left(\frac{\mathbf{P}}{\lambda}\right)^n_{ii}}{\lim_{m\to\infty} \sum_j \left(\frac{\mathbf{P}}{\lambda}\right)^m_{jj}}.$$

To evaluate the numerator and denominator, we observe that **P** is a positive matrix, and therefore, by Perron's theorem (see Golub & Van Loan, 1996), there is a largest positive real eigenvalue, λ, that is, $\lambda > |\mu|$, for all eigenvalues, μ, except $\mu = \lambda$. It follows that

$$\lim_{n\to\infty} \left(\frac{\mathbf{P}}{\lambda}\right)^n = \frac{\mathbf{s}\bar{\mathbf{s}}^\mathbf{T}}{\bar{\mathbf{s}}^\mathbf{T}\mathbf{s}},$$

where $\mathbf{s}$ and $\bar{\mathbf{s}}$ are the right and left eigenvectors of **P** with largest positive real eigenvalue, that is, $\lambda\mathbf{s} = \mathbf{P}\mathbf{s}$ and $\lambda\bar{\mathbf{s}} = \mathbf{P}^\mathbf{T}\bar{\mathbf{s}}$. Because of the time-reversal

symmetry of **P**, the right and left eigenvectors are related by a permutation that exchanges values associated with the same edge but opposite directions: $\bar{s}_i = s_{\bar{i}}$. Applying this result to the denominator of the expression for c_i yields

$$\lim_{m\to\infty} \sum_j \left(\frac{\mathbf{P}}{\lambda}\right)^m_{jj} = \sum_j \frac{(\mathbf{s}\bar{\mathbf{s}}^\mathbf{T})_{jj}}{\bar{\mathbf{s}}^\mathbf{T}\mathbf{s}} = 1.$$

Consequently, the saliency of edge i is given by the numerator:

$$c_i = \frac{(\mathbf{s}\bar{\mathbf{s}}^\mathbf{T})_{ii}}{\mathbf{s}^\mathbf{T}\bar{\mathbf{s}}} = \frac{s_i\bar{s}_i}{\sum_j s_j\bar{s}_j}.$$

3.2 Link Saliency and Markov Chains. Generalizing the notion of edge saliency, it is possible to compute the relative number of contours that begin at edge i, immediately visit edge j, and then eventually return to i:

$$C_{ji} = \lim_{n\to\infty} \frac{P^{(n-1)}(i \mid j)P(j \mid i)}{\sum_k P^{(n)}(k \mid k)}.$$

This quantity, which we term the link saliency, is the relative number of closed contours that visit edges i and j in succession. To solve this expression, we first divide the numerator and denominator by λ^n and then take separate limits in the numerator and the denominator, yielding:

$$C_{ji} = \frac{\lim_{n\to\infty} \left(\frac{\mathbf{P}}{\lambda}\right)^{n-1}_{ij} \left(\frac{\mathbf{P}}{\lambda}\right)_{ji}}{\lim_{m\to\infty} \sum_k \left(\frac{\mathbf{P}}{\lambda}\right)^m_{kk}}.$$

After evaluating the limits as before, we get the following expression for link saliency:

$$C_{ji} = \frac{\bar{s}_j P(j \mid i) s_i}{\lambda \bar{\mathbf{s}}^\mathbf{T}\mathbf{s}}.$$

It easy to verify that the number of closed contours entering edge i equals the number of closed contours leaving edge i:

$$\sum_j C_{ij} = c_i = \sum_k C_{ki}.$$

This establishes that closed contours are conserved at edges. Since closed contours are conserved, it is possible to treat them as Markov chains. By dividing the joint probability of a closed contour visiting edges i and j in succession by the probability of a closed contour visiting i, we get a conditional probability,

$$M(j \mid i) = \frac{C_{ji}}{c_i} = \frac{\bar{s}_j P(j \mid i)}{\lambda \bar{s}_i},$$

equal to the probability that a closed contour will visit edge j given that it has just visited i. Unlike the matrix **P**, the matrix **M** is stochastic, that is,

$\sum_j M_{ji} = 1$, and $\mathbf{c}$ is the eigenvector of $\mathbf{M}$ with eigenvalue equal to one:

$$\mathbf{c} = \mathbf{M}\mathbf{c}.$$

This is consistent with our claim that c_i is the probability of a closed contour through edge i.

3.3 Saliency of Closed Contours. In our approach, the magnitude of the largest positive real eigenvalue of $\mathbf{P}$ is related to the saliency of the most salient closed contour. To develop some intuition for the meaning of the eigenvalue and its relationship to contour saliency, it will be useful to consider an idealized situation. We know from linear algebra that the eigenvalues of $\mathbf{P}$ are solutions to the equation $\det(P-\lambda I) = 0$. Now consider a closed contour, Γ, threading n edges. The probability that a contour will visit edge $\Gamma_{(i+1) \bmod n}$, given that it has just visited edge Γ_i, equals $P(\Gamma_{(i+1) \bmod n} \mid \Gamma_i)$. Assuming that the probability of a contour joining edges Γ_i and Γ_j it is negligible for nonadjacent i and j (i.e., $P_{ji} = P(\Gamma_j \mid \Gamma_i)$ when $j = (i+1) \bmod n$ and $P_{ji} = 0$ otherwise), then:

$$\lambda(\Gamma) = \left(\prod_{i=1}^{n} P(\, \Gamma_{(i+1) \bmod n} \mid \Gamma_i) \right)^{\frac{1}{n}}$$

satisfies $\det(P - \lambda I) = 0$. This is the geometric mean of the transition probabilities in the closed path. Equivalently, minus one times the logarithm of the eigenvalue equals the average transition energy:

$$-\ln \lambda(\Gamma) = -\sum_{i=1}^{n} \ln P(\, \Gamma_{(i+1) \bmod n} \mid \Gamma_i)/n.$$

Because maximizing λ minimizes the average transition energy, there is a close relationship between the most salient closed contour and the minimum mean weight cycle of the directed graph with weight matrix, $-\ln \mathbf{P}$.

Using psychophysical methods, Elder and Zucker (1994) quantified the effect that the distribution and size of boundary gaps have on contour closure. They measured reactiontime in a preattentive search task and found that it was well modeled by the square root of the sum of the squares of the gap lengths. They assumed that reactiontime is inversely related to the degree of contour closure. We note that for stimuli consisting of relatively few edges of negligible length separated by large gaps, λ is maximized when the edges are equidistant. Also, the decrease in saliency due to one large gap is much greater than the decrease due to many small gaps. Both of these properties are consistent with the observations by Elder and Zucker (1994).

3.4 Stochastic Completion Fields. Finally, given $\mathbf{s}$ and $\bar{\mathbf{s}}$, it is possible to compute the relative number of closed contours at an arbitrary position and

direction in the plane, that is, to compute the stochastic completion field. Let $\eta = (\vec{u}, \phi)$ be an arbitrary position and direction in the plane; then

$$c_\eta = \lim_{n\to\infty} \sum_i \sum_j \frac{P^{(n-1)}(i \mid j) P(j \mid \eta) P(\eta \mid i)}{\sum_k P^{(n)}(k \mid k)}$$

represents the probability that a contour first visits edge i, then passes through position $\vec{u}$ in direction ϕ, next visits edge j, then visits additional $n-1$ edges, and finally returns to i. Dividing numerator and denominator by λ^n yields

$$c_\eta = \lim_{n\to\infty} \sum_i \sum_j \left[\left(\frac{\left(\frac{\mathbf{P}}{\lambda}\right)^{n-1}_{ij}}{\sum_k \left(\frac{\mathbf{P}}{\lambda}\right)^n_{kk}} \right) \cdot \left(\frac{P(j \mid \eta) P(\eta \mid i)}{\lambda} \right) \right].$$

Taking separate limits in the numerator and the denominator results in

$$c_\eta = \sum_i \sum_j \left[\left(\frac{s_i \bar{s}_j}{\sum_k s_k \bar{s}_k} \right) \cdot \left(\frac{P(j \mid \eta) P(\eta \mid i)}{\lambda} \right) \right],$$

which can be rearranged to yield

$$c_\eta = \frac{1}{\lambda \mathbf{s}^\mathbf{T} \bar{\mathbf{s}}} \underbrace{\sum_i P(\eta \mid i) s_i}_{\text{source field}} \cdot \underbrace{\sum_j P(j \mid \eta) \bar{s}_j}_{\text{sink field}}.$$

This expression gives the relative probability that a closed contour will pass through η, an arbitrary position and direction in the plane. Note that this is a natural generalization of the factorization of the stochastic completion field into the product of source and sink fields described by Williams and Jacobs (1997a). For this reason, we call the components of $\mathbf{s}$ and $\bar{\mathbf{s}}$ the eigensources and eigensinks of the stochastic completion field. The crucial difference is that we now know how to weight the contribution of each edge to the stochastic completion field.

4 A Continuous-State, Discrete-Time Random Process

The approach just outlined suffers from several limitations. First, it assumes that we have perfect knowledge of the positions and directions of the edges that serve as the input to the problem. This is unecessarily restrictive. Second, because the vectors and matrices (e.g., $\mathbf{s}$ and $\mathbf{P}$) are specific to the configuration of input edges, it is not obvious how the computations we have described can be translated into brainlike representations and algorithms, that is, parallel operations in a finite basis.

In order to address both of these limitations, we can generalize our approach by considering the distribution of closed contours, $\mathbf{c}$, to be a function

of position and direction in the plane, $c(\vec{u}, \phi)$. While previously the input took the form of a set of edges, with exact knowledge of the edge's positions and orientations, now the input takes the form of a probability density function (pdf), $b(\vec{x}, \theta)$, which represents the probability that an edge exists at position $\vec{x}$ and orientation θ. We refer to this pdf as the input bias function. Instead of a discrete-state and discrete-time random process with transition probabilities represented by a matrix, $\mathbf{P}$, we have a continuous-state and discrete-time random process with transition probabilities represented by a linear operator, $P(\vec{u}, \phi \mid \vec{x}, \theta)$. The expression for the eigensources of the stochastic completion field then becomes

$$\lambda s(\vec{x}, \theta) = \int \int \int_{\mathbf{R}^2 \times S^1} d\vec{u} d\phi \; Q(\vec{x}, \theta \mid \vec{u}, \phi) s(\vec{u}, \phi),$$

where

$$Q(\vec{x}, \theta \mid \vec{u}, \phi) = b(\vec{x}, \theta)^{\frac{1}{2}} P(\vec{x}, \theta \mid \vec{u}, \phi) b(\vec{u}, \phi)^{\frac{1}{2}},$$

and $s(\vec{x}, \theta)$ is the eigenfunction of Q with largest positive real eigenvalue λ. The input bias function, $b(.)$, is distributed equally between the left and right sides of $P(.)$ to preserve the time-reversal symmetry of $Q(.)$. Consequently, left and right eigenfunctions of $Q(.)$ with equal eigenvalue are related through a reversal symmetry, $\bar{s}(\vec{x}, \theta) = s(\vec{x}, \theta + \pi)$. Finally, the expression for the stochastic completion field itself can be generalized in the same way:

$$c(\vec{u}, \phi) = \frac{1}{\lambda \langle s, \bar{s} \rangle} \int \int \int_{\mathbf{R}^2 \times S^1} d\vec{x} d\theta \; P(\vec{u}, \phi \mid \vec{x}, \theta) b(\vec{x}, \theta)^{\frac{1}{2}} s(\vec{x}, \theta) \cdot \int \int \int_{\mathbf{R}^2 \times S^1} d\vec{x}\,' d\theta' \; P(\vec{x}\,', \theta' \mid \vec{u}, \phi) b(\vec{x}\,', \theta')^{\frac{1}{2}} \bar{s}(\vec{x}\,', \theta'),$$

where $\bar{s}(\vec{x}, \theta) = s(\vec{x}, \theta + \pi)$, $\mathbf{R}^2 \times S^1$ is the space of positions in the plane and directions on the circle, and $< s, \bar{s} >= \int \int \int_{\mathbf{R}^2 \times S^1} d\vec{x} d\theta \; s(\vec{x}, \theta) \bar{s}(\vec{x}, \theta)$.

Given the above expression for the stochastic completion field, it is clear that the key problem is computing the eigenfunction with the largest positive real eigenvalue. To accomplish this, we can use the well-known power method (see Golub & Van Loan, 1996). In this case, the power method involves repeated application of the linear operator $Q(.)$ to the function $s(.)$, followed by normalization:

$$s^{(n+1)}(\vec{x}, \theta) = \frac{\int \int \int_{\mathbf{R}^2 \times S^1} d\vec{u} d\phi \; Q(\vec{x}, \theta \mid \vec{u}, \phi) s^{(n)}(\vec{u}, \phi)}{\int \int \int_{\mathbf{R}^2 \times S^1} d\vec{u} d\phi \; s^{(n)}(\vec{u}, \phi)}.$$

In the limit, as n gets very large, $s^{(n+1)}(\vec{x}, \theta)$ converges to the eigenfunction of $Q(.)$, with largest positive real eigenvalue. We observe that the above computation can be considered a continuous-state, discrete-time, recurrent neural network.

5 Scale Invariance

Ideally, we would like our computation to be scale invariant. A computation is scale invariant if scaling the input by a constant factor γ produces a corresponding scaling of the output. This property is best summarized by a commutative diagram:

$$\begin{array}{ccc} b(\vec{x},\theta) & \stackrel{C}{\rightarrow} & c(\vec{x},\theta) \\ \downarrow S & & \downarrow S \\ b(\vec{x}/\gamma,\theta) & \stackrel{C}{\rightarrow} & c(\vec{x}/\gamma,\theta), \end{array}$$

where $b(.)$ is the input, $c(.)$, is the output, $\stackrel{C}{\rightarrow}$ is the computation, and $\stackrel{S}{\rightarrow}$ is the scaling operator. The diagram shows that the output is independent of the order in which the operators are applied.

The only scale-dependent parameter in the computation is the speed of the particles. The lack of scale invariance is due to the fact that this parameter has been arbitrarily set to one. In order to achieve a scale-invariant computation, we need to eliminate this bias. To accomplish this, all speeds must be treated uniformly.

Previously $Q(.)$ was indexed by four arguments: the initial and final particle positions and directions. In the scale-invariant computation, $Q(.)$ will be indexed by six arguments: the initial and final particle positions, directions, and speeds. Because particles have constant speed, $Q(.)$ is block diagonal. This property of $Q(.)$, together with the scale invariance of the Green's function, $G(.)$, allows $Q(.)$ to be defined as follows:

$$Q(\vec{x},\theta,\gamma_1 \mid \vec{u},\phi,\gamma_0) = \begin{cases} b^{\frac{1}{2}}(\vec{x},\theta)P(\vec{x}/\gamma_1,\theta \mid \vec{u}/\gamma_0,\phi)b^{\frac{1}{2}}(\vec{u},\phi) & \text{if } \gamma_0 = \gamma_1 \\ 0 & \text{otherwise.} \end{cases}$$

The $Q(.)$ operator now includes an integral over all positive speeds, γ_0. This eliminates the scale dependency in the computation:

$$\lambda s(\vec{x},\theta,\gamma_1) = \int\int\int_{\mathbf{R}^2\times S^1} d\vec{u}d\phi \int_{\mathbf{R}>0} d\gamma_0\ Q(\vec{x},\theta,\gamma_1 \mid \vec{u},\phi,\gamma_0)s(\vec{u},\phi,\gamma_0).$$

The eigenfunction of $Q(.)$ with the largest positive real eigenvalue represents the limiting distribution for particles of all speeds. Because $Q(.)$ is block diagonal, its eigenfunctions are zero everywhere outside a single region of constant speed. Consequently, the eigenfunction with the largest positive real eigenvalue of $Q(.)$ can be identified in two steps. First, we find the largest positive real eigenvalue for each constant speed submatrix:

$$\lambda(\gamma) = \max \frac{\int\int\int_{\mathbf{R}^2\times S^1} d\vec{x}d\theta\ \bar{s}(\vec{x},\theta) \int\int\int_{\mathbf{R}^2\times S^1} d\vec{u}d\phi\ Q(\vec{x},\theta,\gamma \mid \vec{u},\phi,\gamma)s(\vec{u},\phi)}{\int\int\int_{\mathbf{R}^2\times S^1} d\vec{x}d\theta\ \bar{s}(\vec{x},\theta)s(\vec{x},\theta)},$$

where the eigenvalue is written as a Rayleigh quotient, and the maximum is taken over all eigenfunctions $s(.)$ of the submatrix of $Q(.)$ with constant speed, γ. Next, we find the speed, $\gamma_{\max}$, which maximizes $\lambda(\gamma)$:

$$\gamma_{\max} = \operatorname*{argmax}_{\gamma} \lambda(\gamma).$$

The eigenfunction $s(\vec{x}, \theta)$ with eigenvalue $\lambda(\gamma_{\max})$ gives the limiting distribution for particles of all speeds:

$$\lambda(\gamma_{\max}) s(\vec{x}, \theta) = \int \int \int_{\mathbf{R}^2 \times S^1} d\vec{u} d\phi \; Q(\vec{x}, \theta, \gamma_{\max} \mid \vec{u}, \phi, \gamma_{\max}) s(\vec{u}, \phi).$$

6 Orientation Selectivity in Primary Visual Cortex

A long-standing problem in visual neuroscience is the emergence of orientation-selective responses in simple cells of primary visual cortex given input from cells in lateral geniculate that exhibit little or no orientation selectivity. In contrast with the classical pure feedforward model for the emergence of orientation selectivity proposed by Hubel and Wiesel (1962), the authors of a recent review article (Sompolinsky & Shapley, 1997) and two computer simulation studies (Somers, Nelson, & Sur, 1995); and (Ben-Yishai, Lev Bar-Or, & Sompolinsky, 1995), argue for a model with three defining features: (1) a weak orientation bias provided by excitatory input from the lateral geniculate, (2) intracortical excitatory connections between simple cells with similar orientation preferences, and (3) intracortical inhibitory connections between simple cells without regard to orientation preference. These authors suggest that the weak orientation bias provided by excitatory input from the lateral geniculate is amplified by intracortical excitatory connections between simple cells with similar orientation preference. The role of the intracortical inhibition is to prevent the level of activity due to the intracortical excitation from growing unbounded.

The principal contribution of these recent studies is a unified explanation of the many different (and sometimes contradictory) experimental findings related to the emergence of orientation selectivity in primary visual cortex. However, none of these authors considers the functional significance of orientation selectivity—what purpose it serves in the larger context of human visual information processing. Stated differently, is orientation selectivity an end in itself? Or can it be understood only as an emergent property of a higher-level visual computation, such as contour completion?

To our knowledge, the first model of orientation selectivity in visual cortex that differed significantly from the original Hubel and Wiesel feedforward model was described in Parent and Zucker (1989). Like the more recent and detailed integrate-and-fire models described in Somers et al. (1995) and Ben-Yishai et al. (1995), Parent and Zucker considered orientation selectivity to be an end in itself. Unlike these recent models, Parent

and Zucker's primary motivations were computational. In the relaxation labeling network they describe, crude local estimates of tangent and curvature (i.e., the initial states of simple cells with and without endstopping) are sharpened by the activity of recurrent excitatory connections representing geometric constraints between position, tangent, and curvature. The magnitude of the network state vector is normalized at every time step by dividing by the sum of its components.[1] It can be seen that this divisive normalization plays the same role as the nonspecific inhibition in the model of Somers et al. (1995). Consequently, we see that there is a strong relationship between Parent and Zucker's model and more recent models of orientation selectivity in primary visual cortex.

In this article, we do not consider orientation selectivity to be an end in itself; rather, we consider it to be an emergent property of a higher-level visual computation devoted to contour completion. Unlike Parent and Zucker (1989), we did not specifically intend to model the emergence of orientation selectivity in primary visual cortex. Instead, our intention was to formulate a computational theory level (Marr, 1982) account of contour completion; orientation selectivity is simply a sideeffect. Our specific hypothesis is that one of the major goals of early visual processing is to compute a scale-invariant distribution of closed contours, $c(.)$, consistent with weak constraints on position and direction derived by linear filtering, $b(.)$. We termed this distribution the stochastic completion field and in the previous section described a continuous-state, discrete-time neural network for computing it. We have demonstrated experimentally (see the next section) that the distribution of $s(.)$, the eigensources of $c(.)$, can be highly nonisotropic, even for isotropic $b(.)$. We now show that the neural network that computes $s(.)$ is consistent with recent hypotheses concerning the emergence of orientation selectivity in primary visual cortex.

The state of the neural network at time t is given by $s^{(t)}(.)$, which is a function of $\mathbf{R}^2 \times S^1$, the continuous space of positions and directions. In this article, we do not address the problem of how $s^{(t)}(.)$ can be represented as a weighted sum of a fixed set of basis functions, that is, receptive fields. Obviously this needs to be done before we can claim to have a complete account of the computation at the algorithm and representation level (Marr, 1982).[2] Nevertheless, we believe that it is often best first to describe the

[1] For an introduction to relaxation labeling, see Rosenfeld, Hummel, and Zucker (1976).

[2] Although beyond the scope of this article, the problem of representing continuous (but band-limited) functions of position and direction using a finite set of basis functions is one we are actively working on. For example, in Zweck and Williams (2000), we consider the problem of computing the stochastic completion field using parallel operations in a finite basis subject to the constraint that the result be invariant under rotations and translations of the input pattern. This is accomplished, in part, by generalizing the notions of steerability and shiftability of basis functions introduced in Simoncelli, Freeman, Adelson, and Heeger (1992).

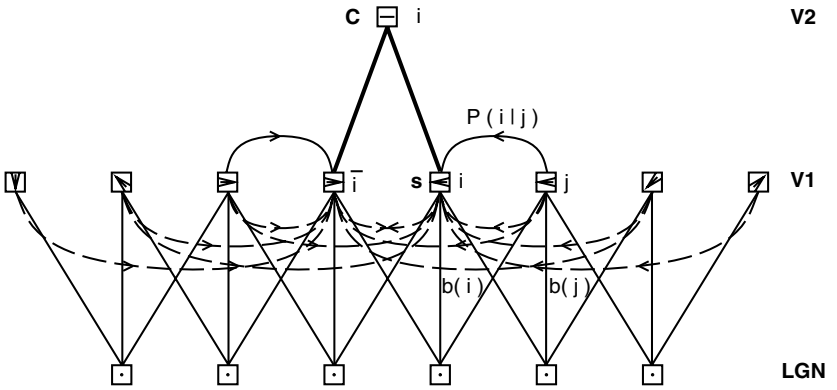

Figure 3: Thin solid lines indicate feedforward connections from LGN, which provide a weak orientation bias, that is, $b(i)$, to simple cells in V1, that is, $s(i)$. Solid lines with arrows indicate orientation-specific intracortical excitatory connections—$P(i \mid j)$. Dashed lines with arrows indicate orientation-nonspecific intracortical inhibitory connections. Thick solid lines indicate feedforward connections between V1 and V2, that is, $c(i)$.

computation in the continuum (e.g., Williams & Jacobs, 1997a), and by doing so temporarily avoid the issue of sampling altogether.[3]

Recall that the fixed point of the neural network we described is the eigenfunction with the largest positive real eigenvalue of the linear operator $Q(.)$. The linear operator $Q(.)$ is the composition of the input-independent linear operator $P(.)$ and the input-dependent linear operator $B(.)$. The dynamics of the neural network are derived from the update equation for the standard power method for computing eigenvectors. It is useful to draw an analogy between our neural network for contour completion and the models for the emergence of orientation selectivity in primary visual cortex described by Somers et al. (1995) and Ben-Yishai et al. (1995). (See Figure 3.) First, we can identify $s(.)$ with simple cells in V1 and the input bias function $b(.)$, which modulates $s(.)$ in the numerator of the update equation, with the feedforward excitatory connections from the lateral geniculate. Second, we can identify $P(.)$ with the intracortical excitatory connections that Somers et al. hypothesize are primarily responsible for the emergence of orientation selectivity in V1. As in the model of Somers et al., these connections are highly specific and mainly target cells of similar orientation

[3] In this respect, the model of Parent and Zucker (1989) is instructive. Although their model was a major source of inspiration for our own, we believe that it (unnecessarily) confounds the computational theory level goal of estimating tangent and curvature everywhere with algorithm and representation level details related to discrete sampling.

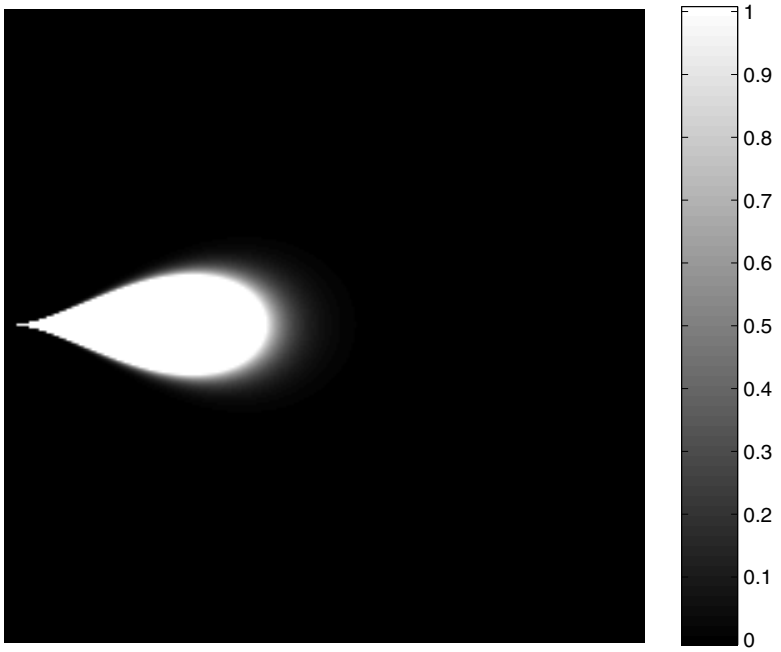

Figure 4: Visualization of $\int_{S^1} d\theta P(x, y, \theta \mid 0, 0, 0)$ computed using the analytic expression from Thornber and Williams (1996). This is a rendering of the kernel of the hypothesized intracortical excitatory connections, integrated over the θ dimension. Displayed values are scaled by a factor of 10^5.

preference (see Figures 4 and 5). Third, we identify the denominator of the update equation with the non-specific intra-cortical inhibitory connections that Somers et al. hypothesize keep the level of activity within bounds. Because the purpose of the denominator is to normalize $s(.)$, it plays the same role as the denominator in the relaxation labeling update equation (Rosenfeld et al., 1976) and might be implemented using a mechanism similiar to the divisive inhibition mechanism proposed by Heeger (1992). Finally, we identify $c(.)$, the stochastic completion field, with the population of cells in V2 described by von der Heydt, Peterhans, and Baumgartner (1984).

There is obviously a huge gap in level of detail between the continuous-state, discrete-time neural network we describe and the integrate-and-fire simulations of Somers et al. (1995) and Ben-Yishai et al. (1995). For this reason, the above discussion must be regarded as highly speculative. However, we would like to stress that unlike these recent theoretical studies, our neural network implements a well-defined (and nontrivial) computation in the

Figure 5: Visualization of $\int_{\mathbf{R}} dx P(x, y, \theta \mid 0, 0, 0)$ computed using the analytic expression from Thornber and Williams (1996). This is a rendering of the kernel of the hypothesized intracortical excitatory connections, integrated over the x dimension. Displayed values are scaled by a factor of 10^5.

sense of Marr (1982). For this reason, we believe our top-down approach complements the bottom-up approach pursued by others.

7 Experiments

7.1 Analytic Solution of Conditional Probabilities. The conditional probability, $P(i \mid j)$, is the probability that a particle, moving with constant speed in a direction given by a Brownian motion, will travel from edge i to edge j. In order to test the computational theory described in the previous sections, we need a fast and efficient method for computing these probabilities. In prior work, these probabilities were computed using Monte Carlo simulation (Williams & Jacobs, 1997a) and by numerical solution of the Fokker-Planck equation (Williams & Jacobs, 1997b). Although there (currently) is no analytic solution for the Fokker-Planck equation described by Mumford (1994), there is a similiar equation for which an exact analytic solution exists (Thornber & Williams, 1996). This equation governs the motion of particles with position and velocity (instead of position and direction). Straight lines are base trajectory of these particles. Their velocities are modified by random impulses with a zero-mean distribution, and variance, σ_g^2, acting at Poisson distributed times, with rate, R_g. If the initial and final velocities are conditioned to be equal, this random process can be used to compute stochastic completion fields that are virtually indistinguishable from those computed using the Mumford random process. The analytic expression for $P(i \mid j)$, based on the process described in Thornber and Williams (1996), is given in the appendix.

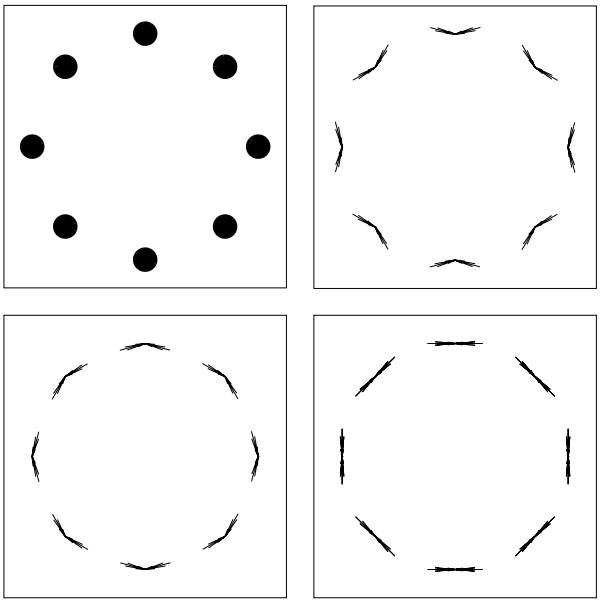

Figure 6: (Top left) The eight position constraints (the dots) that define the test configuration. The order of traversal, directions, and speed are specified a priori. (Top right) The right eigenvector, $\mathbf{s}(\gamma_{\max})$ [where $\gamma_{\max} = 0.149$, represents the limiting distribution of the random process over all spatial scales. (Bottom left) The left eigenvector, $\bar{\mathbf{s}}(\gamma_{\max})$, represents the time-reversed distribution. (Bottom right) The vector, $\mathbf{s}(\gamma_{\max})\ \bar{\mathbf{s}}(\gamma_{\max})$, represents the magnitude of the stochastic completion field at the locations of the dots.

7.2 Eight Dot Circle. Given eight dots spaced uniformly around the perimeter of a circle of diameter $d = 16$, we would like to find the relative number of closed contours that visit each dot and in each direction. We would also like to compute the corresponding completion field (see Figure 6, top left). Neither the order of traversal, directions θ, nor speed γ is specified a priori. Accordingly, the position and direction bias, b_{dot}, is purely isotropic:

$$b_{dot}(\vec{x}, \theta) = \sum_i \delta(\vec{x} - \vec{x}_i).$$

The isotropy of $b_{dot}(\vec{x}, \theta)$ can be verified by noting the lack of θ dependence on the right side of the equation. In our neural model, this represents the assumption that the LGN provides no information about orientation to simple cells in V1.

So that all computations can be performed using ordinary vectors and matrices, the functions $s(.)$, $P(.)$, and $b(.)$ are sampled at the locations of the eight dots, $\vec{x}_i$, and at N discrete directions in the θ dimension, to form a vector, $\mathbf{s}$, and matrices, $\mathbf{P}$ and $\mathbf{B}$:

$$\begin{aligned} s_k &= s(\vec{x}_i, m\Delta_\theta) \\ P_{kl} &= P(\vec{x}_i, m\Delta_\theta \mid \vec{x}_j, n\Delta_\theta) \\ B_{kl} &= \delta_{k\ell} \end{aligned}$$

where $k = iN + m$ and $l = jN + n$, for dots i and j and sampling directions m and n. Since $b(.)$ is isotropic and unweighted, after sampling, $\mathbf{B} = \mathbf{I}$. In all of our experiments, we sample the θ dimension at 5 degree intervals. Consequently, there are $N = 72$ discrete directions and 576 position-direction pairs, that is, $\mathbf{P}$ is of size 576×576.[4] To achieve scale invariance, we make $\mathbf{s}$ and $\mathbf{P}$ functions of γ and solve

$$\lambda(\gamma)\,\mathbf{s}(\gamma) \;=\; \mathbf{B}^{\frac{1}{2}}\mathbf{P}(\gamma)\mathbf{B}^{\frac{1}{2}}\mathbf{s}(\gamma) \;=\; \mathbf{P}\mathbf{s}(\gamma).$$

In the first experiment, we evaluated $\lambda(\gamma)$ over the speed interval $[1.1^{-1}, 1.1^{-30}]$ using standard numerical routines and plotted the magnitude of the largest, real positive eigenvalue, λ, versus $\log_{1.1}(1/\gamma)$ (see Figure 7). The function reaches its maximum value at $\gamma_{\max} \approx 1.1^{-20}$. Consequently, the eigenvector $\mathbf{s}(1.1^{-20})$ represents the limiting distribution over all spatial scales (See Figure 6, top right). The direction-reversed permutation of this eigenvector $\bar{\mathbf{s}}(1.1^{-20})$ is shown in Figure 6 (bottom left). This is the eigenvector of $\mathbf{P}^{\mathbf{T}}$ with eigenvalue $\lambda(\gamma_{\max})$. The component-wise product of $\mathbf{s}$ and $\bar{\mathbf{s}}$ is shown in Figure 6 (bottom right). This vector represents the magnitude of the stochastic completion field at the locations of the dots. As one would expect, orientations tangential to the circle have the greatest magnitude. The magnitude of the stochastic completion field at all other positions in the plane (summed over all directions) is shown in Figure 8.

Next, we scaled the test figure by a factor of two—$d' = 32.0$—and plotted $\lambda'(\log_{1.1}(1/\gamma)$ over the same interval (see Figure 7). We observe that $\lambda'(1.1^{-x+7}) \approx \lambda(1.1^{-x})$; that is, when plotted using a logarithmic x-axis, the functions are identical except for a translation. It follows that $\gamma'_{\max} \approx \log_{1.1} 7 \times \gamma_{\max} \approx 2.0 \times \gamma_{\max}$. This confirms the scale invariance of the system: doubling the size of the figure results in a doubling of the selected speed.

[4] The parameters defining the distribution of completion shapes are $T = 0.0005$ and $\tau = 9.5$. See Thornber and Williams (1996).

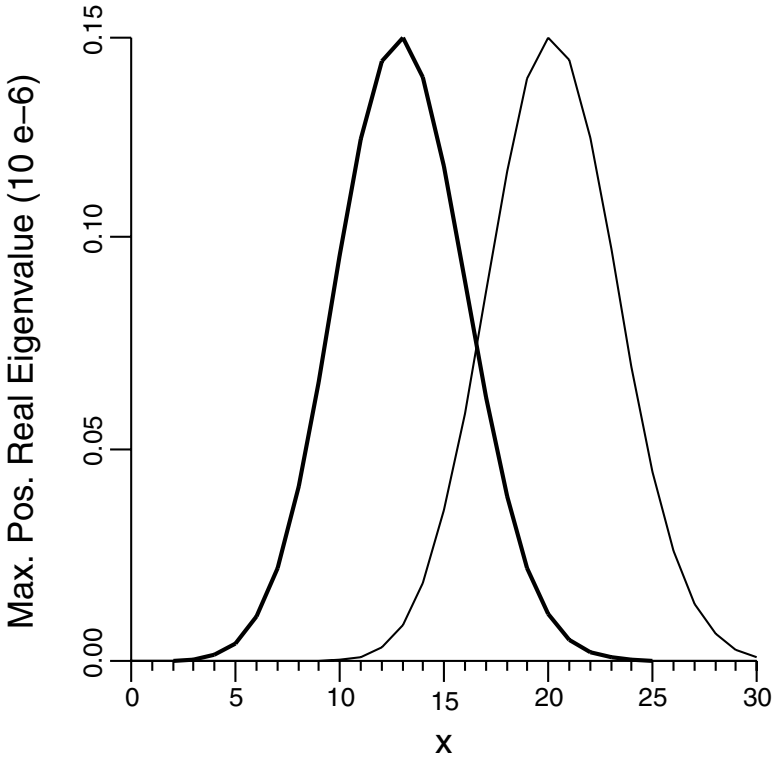

Figure 7: Plot of magnitude of maximum positive real eigenvalue, λ, versus $x = \log_{1.1}(1/\gamma)$ for eight-point test configuration with $d = 16.0$ (thin line) and $d = 32.0$ (thick line). Note that speed increases right to left.

7.3 Contours with Corners. The distribution of shapes considered by Mumford (1994) and Thornber and Williams (1996) consists of smooth, short contours. Yet there are many examples in human vision where completion shapes perceived by humans contain discontinuities in orientation (corners). Figure 9 shows a display by Kanizsa (1979). This display illustrates the completion of a circle and square under a square occluder. The completion of the square is significant because it includes a discontinuity in orientation. Figure 10 shows a continuum of Koffka crosses. When the width of the arms of the cross is increased, observers report that the percept changes from an illusory circle to an illusory square (Sambin, 1974).

In the experiments described in the next section, we did not generalize the distribution described in Mumford (1994) to include contours with corners (e.g., by randomizing the direction of the particle's motion at Poisson distributed times). Instead, consistent with the experiments in the previ-

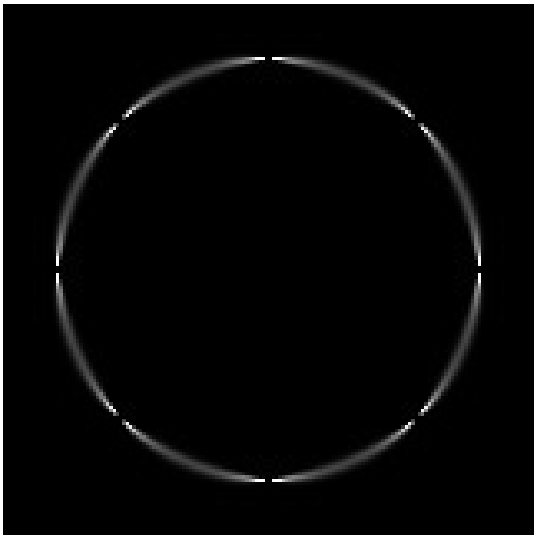

Figure 8: Stochastic completion field due to $\mathbf{s}(\gamma_{\max})$ for eight-point circle.

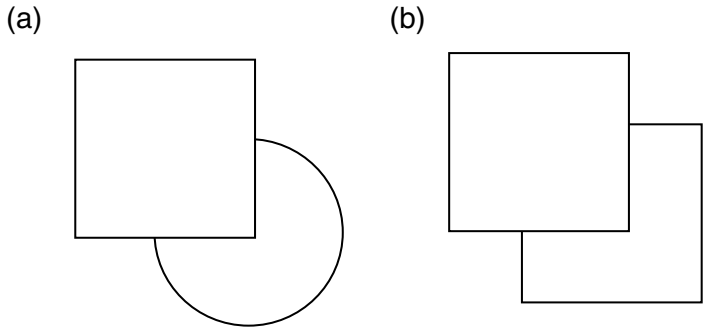

Figure 9: Amodal completion of a partially occluded circle and square (redrawn from Kanizsa, 1979).

ous section, we assumed a distribution of completion shapes consisting of straight-line base trajectories modified by random impulses drawn from a mixture of two limiting distributions. The first distribution consists of weak but frequently acting impulses (we call this the gaussian limit). The distribution of these weak impulses has zero mean and variance equal to σ_g^2. The weak impulses act at Poisson times with rate R_g. The second distribution consists of strong but infrequently acting impulses (we call this the Poisson limit). Here, the magnitude of the random impulses is gaussian distributed

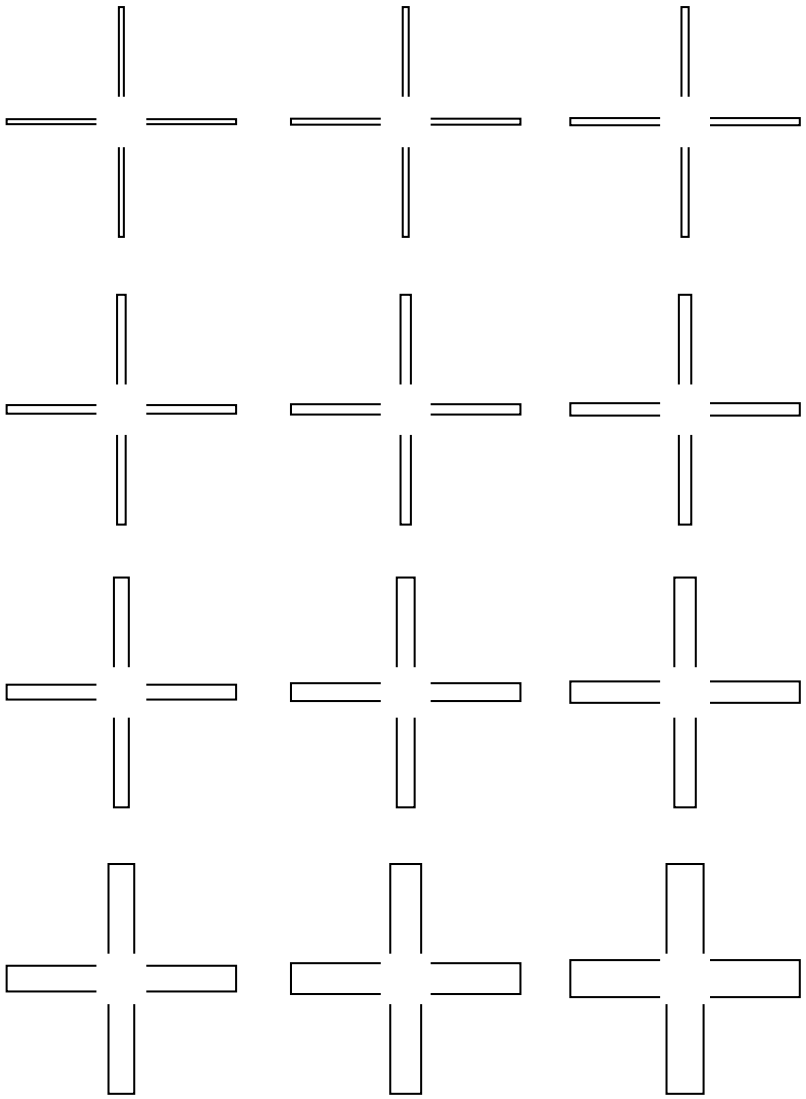

Figure 10: An array of Koffka crosses with arms of varying width. Observers report that as the width of the arms increases, the shape of the illusory contour changes from a circle to a square.

with zero mean. However, the variance is equal to σ_p^2 (where $\sigma_p^2 \gg \sigma_g^2$). The strong impulses act at Poisson times with rate $R_p \ll R_g$. Particles decay with half-life equal to a parameter τ. The effect is that particles tend

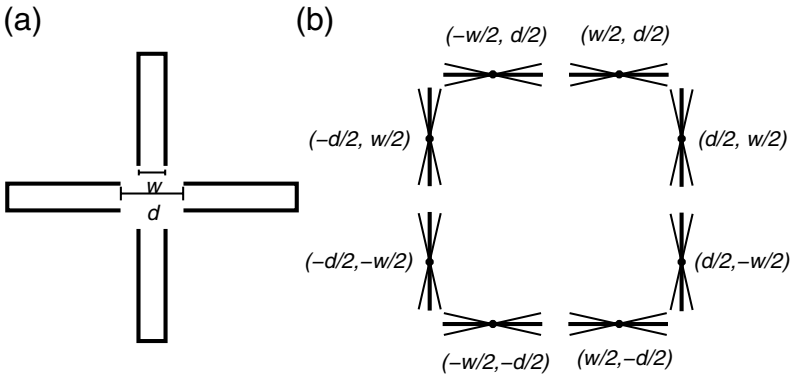

Figure 11: (a) Koffka cross. (b) Orientation and position constraints in terms of d and w. The normal orientation at each end point is indicated by the thick line; the thin lines represent plus or minus one standard deviation (i.e., 12.8 degrees) of the gaussian weighting function.

to travel in smooth, short paths punctuated by occasional orientation discontinuities. (The interested reader is encouraged to consult Thornber and Williams, 2000.)

7.4 Koffka Cross. The Koffka cross stimulus (see Figure 10) has two basic degrees of freedom that we call diameter (d) and arm width (w) (see Figure 11 a). We are interested in how the stochastic completion field changes as these parameters are varied. (Recall that observers report that as the width of the arms increases, the shape of the illusory contour changes from a circle to a square; Sambin, 1974.) The end points of the lines comprising the Koffka cross can be used to define a set of position and orientation constraints (see Figure 11 b). The position constraints are specified in terms of the parameters d and w. The orientation constraints take the form of a gaussian weighting function that assigns higher probabilities to contours passing through the end points with orientations normal to the lines. Observe that Figure 12a is perceived as a square while Figure 12b is perceived as a circle. Yet the positions of the line end points are the same. It follows that the orientations of the lines affect the percept. We have chosen to model this dependence through the use of a gaussian weighting function that favors contours passing through the end points of the lines in the normal direction. The corresponding input bias function is

$$b_{end}(\vec{x},\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \sum_i \delta(\vec{x}-\vec{x}_i) e^{-(\theta-\theta_i \pm \frac{\pi}{2})/2\sigma^2},$$

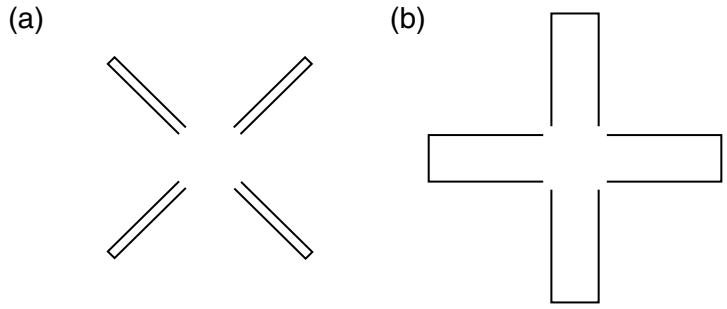

Figure 12: (a) Typically perceived as circle. (b) Typically perceived as square. The positions of the ends of the line segments are the same in both cases.

where $\sigma = 12.8$ degrees is the standard deviation of the gaussian weighting function. As before, so that all computations can be performed using ordinary vectors and matrices, the functions $s(.)$, $P(.)$, and $b(.)$ are sampled at the locations of the eight line end points $\vec{x}_i$ and at N discrete directions in the θ dimension. The vector $\mathbf{s}$ and matrix $\mathbf{P}$ are defined as before. However, since $b(.)$ is not isotropic, $\mathbf{B}$ is now a diagonal matrix,

$$B_{kk} = b_{end}(\vec{x}_i, n\Delta_\theta),$$

where $k = iN + n$, for line end point i and sampling direction n. To achieve scale invariance, we make $\mathbf{s}$ and $\mathbf{P}$ functions of γ and solve

$$\lambda(\gamma)\,\mathbf{s}(\gamma) \;=\; \mathbf{B}^{\frac{1}{2}}\mathbf{P}(\gamma)\mathbf{B}^{\frac{1}{2}}\mathbf{s}(\gamma) \;=\; \mathbf{Q}(\gamma)\mathbf{s}(\gamma),$$

where $\mathbf{P}(\gamma)$ is the edge-to-edge transition probability matrix for speed γ, $\lambda(\gamma)$ is an eigenvalue of $\mathbf{Q}(\gamma)$, and $\mathbf{s}(\gamma)$ is the corresponding eigenvector. Let $\lambda(\gamma)$ be the largest positive real eigenvalue of $\mathbf{Q}(\gamma)$, and let $\gamma_{\max}$ be the scale where $\lambda(\gamma)$ is maximized. Then $\mathbf{s}(\gamma_{\max})$, the eigenvector of $\mathbf{Q}(\gamma_{\max})$ associated with $\lambda(\gamma_{max})$, is the limiting distribution over all spatial scales.

First, we used a Koffka cross where $d = 2.0$ and $w = 0.5$ and evaluated $\lambda(\gamma)$ over the speed interval $[8.0\times1.1^{-1}, 8.0\times1.1^{-80}]$ using standard numerical routines.[5] The function reaches its maximum value at $\gamma_{\max} \approx 8.0\times1.1^{-62}$ (see Figure 13). Observe that the completion field due to the eigenvector, $\mathbf{s}(8.0 \times 1.1^{-62})$ is dominated by contours of a predominantly circular shape (see Figure 14, right). We then uniformly scaled the Koffka cross figure by a factor of two—$d' = 4.0$ and $w' = 1.0$—and plotted $\lambda'(\log_{1.1} 1/\gamma)$ over the

[5] The parameters defining the distribution of completion shapes are $T = 0.0005$, $\tau = 9.5$, $\xi_p = 100.0$, and $R_p = 1.0 \times 10^{-8}$. See Thornber and Williams (2000). As an anti-aliasing measure, the transition probabilities, $P(j \mid i)$, were averaged over initial conditions modeled as gaussians of variance $\sigma_x^2 = \sigma_y^2 = 0.00024$ and $\sigma_\theta^2 = 0.0019$.

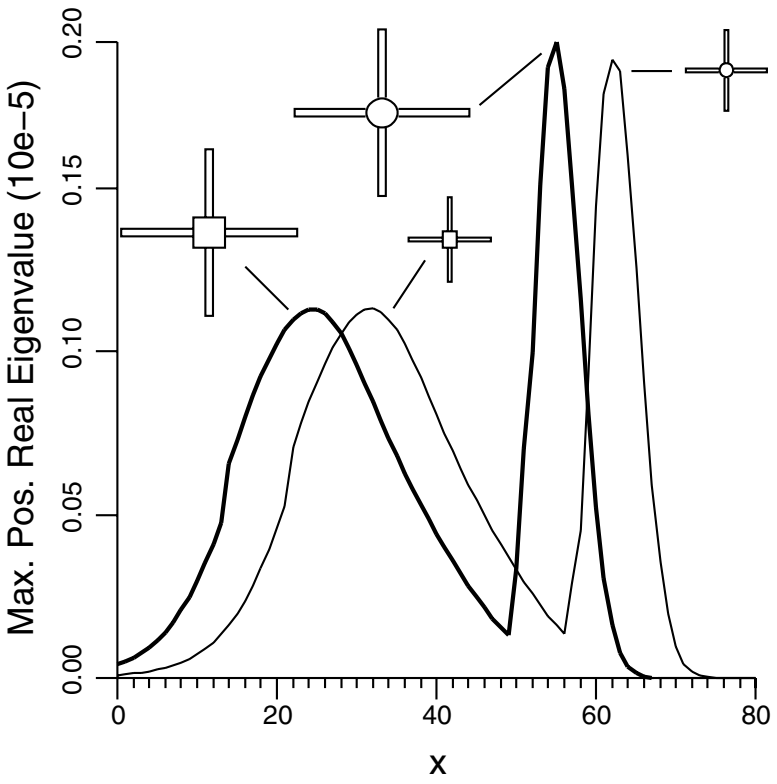

Figure 13: Plot of magnitude of maximum positive real eigenvalue λ versus $x = \log_{1.1}(1/\gamma)$ for Koffka crosses with $d = 2.0$ and $w = 0.5$ (thin line) and $d = 4.0$ and $w = 1.0$ (thick line).

same interval (see Figure 13). Observe that $\lambda'(8.0 \times 1.1^{-x+7}) \approx \lambda(8.0 \times 1.1^{-x})$. As before, this confirms the scale invariance of the system.

Next, we studied how the relative magnitudes of the local maxima of $\lambda(\gamma)$ change as the parameter w is varied. We begin with a Koffka Cross where $d = 2.0$ and $w = 0.5$ and observe that $\lambda(\gamma)$ has two local maxima (see Figure 15). We refer to the larger of these maxima as γ_{circle}. As previously noted, this maximum is located at approximately 8.0×1.1^{-62}. The second maximum is located at approximately 8.0×1.1^{-32}. When the completion field due to the eigenvector $\mathbf{s}(8.0 \times 1.1^{-32})$ is rendered, we observe that the distribution is dominated by contours of predominantly square shape (see Figure 16a). For this reason, we refer to this local maximum as γ_{square}. Now consider a Koffka cross where the widths of the arms are doubled but the diameter remains the same: $d' = 2.0$ and $w' = 1.0$. We observe that $\lambda'(\gamma)$ still

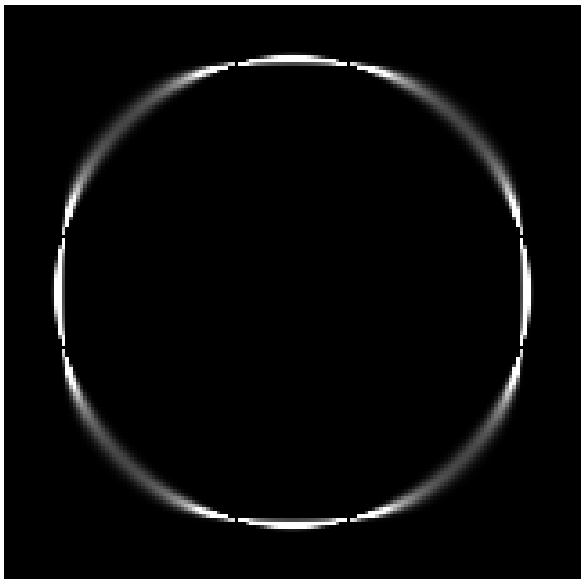

Figure 14: Stochastic completion field due to the eigenvector $\mathbf{s}(8.0 \times 1.1^{-62})$. This is the eigenvector with maximum positive real eigenvalue for a Koffka cross with $d = 2.0$ and $w = 0.5$.

has two local maxima: one at approximately 8.0×1.1^{-63} and a second at approximately 8.0×1.1^{-29} (see Figure 15). When we render the completion fields due to the eigenvectors $\mathbf{s}'(8.0 \times 1.1^{-63})$ and $\mathbf{s}'(8.0 \times 1.1^{-29})$, we find that the completion fields have the same general character as before: the contours associated with the smaller spatial scale (i.e., lower speed) are approximately circular, and those associated with the larger spatial scale (i.e., higher speed) are approximately square (see the lower right and lower left of Figure 16). Accordingly, we refer to the locations of the respective local maxima as γ'_{circle} and γ'_{square}. However, what is most interesting is that the relative magnitudes of the local maxima have reversed. Whereas we previously observed that $\lambda(\gamma_{circle}) > \lambda(\gamma_{square})$, we now observe that $\lambda'(\gamma'_{square}) > \lambda'(\gamma'_{circle})$. Therefore, the completion field due to the eigenvector $\mathbf{s}'(\gamma'_{square})$ [not $\mathbf{s}'(\gamma'_{circle})$!] represents the limiting distribution over all spatial scales. This is consistent with the transition from circle to square reported by human observers when the widths of the arms of the Koffka cross are increased.

8 Conclusion

We have improved on a previous model of illusory contour formation by showing how to compute a scale-invariant distribution of closed contours

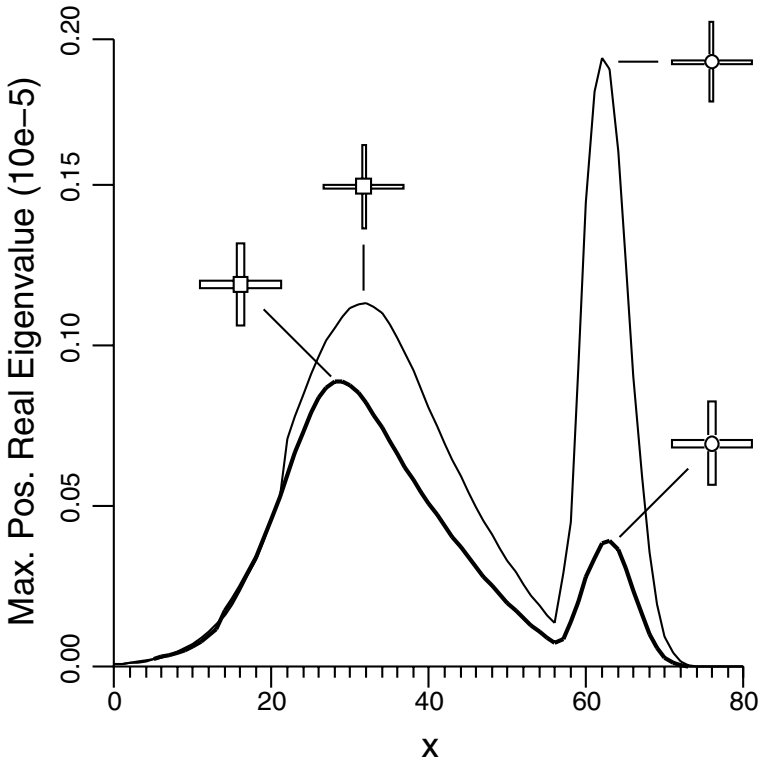

Figure 15: Plot of magnitude of maximum positve real eigenvalue λ versus $x = \log_{1.1}(1/\gamma)$ for Koffka crosses with $d = 2.0$ and $w = 0.5$ (thin line) and $d = 2.0$ and $w = 1.0$ (thick line).

given position constraints alone. We also used our model to explain a previously unexplained perceptual effect.

Appendix

In this appendix, we give the analytic expression for the conditional probabilities in the pure gaussian case.[6] We define the affinity P_{ji} between two directed edges i and j, to be

$$P_{ji} \equiv P(j \mid i) = \int_0^\infty dt\, P(j \mid i; t) \approx F\, P(j \mid i; t_{opt}),$$

[6] For a derivation of a related affinity function, see Sharon, Brandt, and Basri (1997).

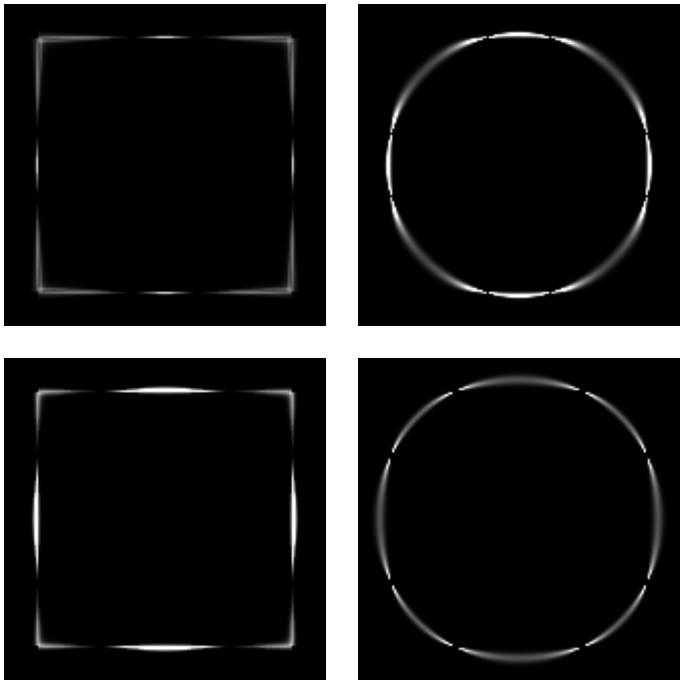

Figure 16: Stochastic completion fields for Koffka cross due to (upper left) $\mathbf{s}(\gamma_{square})$ is a local optimum for $w = 0.5$, (upper right) $\mathbf{s}(\gamma_{circle})$ is the global optimum for $w = 0.5$, (lower left) $\mathbf{s}'(\gamma'_{square})$ is the global optimum for $w = 1.0$, and (lower right) $\mathbf{s}'(\gamma'_{square})$ is a local optimum for $w = 1.0$. These results are consistent with the circle-to-square transition perceived by human subjects when the width of the arms of the Koffka cross is increased.

where $P(j \mid i; t)$ is the probability that a particle that begins its stochastic motion at $(\vec{x}_i, \theta_i)$ at time 0 will be at $(\vec{x}_j, \theta_j)$ 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_{opt}), and a weighting factor, F. The expression for P at time t is

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

where

$$a = [2 + \cos(\theta_j - \theta_i)]\,/\,3$$

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

for $x_{ji} = x_j - x_i$ and $y_{ji} = y_j - y_i$. The parameters T, τ, and γ determine the distribution of shapes (where $T = \sigma_g^2$ is the diffusion coefficient, τ is particle half-life, and γ is speed). The expression for P should be evaluated at $t = t_{opt}$, where t_{opt} is real and positive and satisfies the following cubic equation:

$$-7\,t^3/4 + 3(a\,t^2 - 2\,b\,t + 3\,c)/T = 0.$$

If more than one real, positive root exists, then the root-maximizing $P(j \mid i; t)$ is chosen.[7] Finally, the weighting factor F is

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

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

References

Ben-Yishai, R., Lev Bar-Or, R., & Sompolinsky, H. (1995). Theory of orientation tuning in visual cortex. *Proc. Natl. Acad. of Sci. (USA), 92*, 3844–3848.

Elder, J. H., & Zucker, S. W. (1994). A measure of closure. *Vision Research, 34*(24), 3361–3370.

Golub, G. H., & Van Loan, C. F. (1996). *Matrix computations*. Baltimore, MD: Johns Hopkins University Press.

Heeger, D. J. (1992). Normalization of cell responses in cat striate cortex. *Visual Neuroscience, 9*, 181–197.

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture of the cat's visual cortex. *Journal of Physiology (London), 160*, 106–154.

Kanizsa, G. (1979). *Organization in vision*. New York: Praeger.

Kovacs, I., & 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.

Marr, D. (1982). *Vision*. San Francisco: W. H. Freeman.

Mumford, D. (1994). Elastica and computer vision. In C. Bajaj (Ed.), *Algebraic geometry and its applications*. New York: Springer-Verlag.

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

[7] For a discussion on solving cubic equations, see Press, Flannery, Teukolsky, and Vetterling (1988).

Press, W. H., Flannery, B. P., Teukolsky, S. A., & Vetterling, W. T. (1988). *Numerical recipes in C*. Cambridge: Cambridge University Press.

Rosenfeld, A., Hummel R., & S. Zucker, Scene Labeling by Relaxation Operations, *IEEE Trans. on Systems, Man and Cybernetics, 6*, pp. 420–433, 1976.

Sambin, M. (1974). Angular margins without gradients. *Italian Journal of Psychology, 1*, 355–361.

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

Simoncelli, E., Freeman, W., Adelson E., & Heeger, D., (1992). Shiftable multiscale transforms. *IEEE Trans. Information Theory, 38*(2), 587–607.

Somers, D. C., Nelson, S. B., & Sur, M. (1995). An emergent model of orientation selectivity in cat visual cortical cells. *Journal of Neuroscience, 15*, 5448–5465.

Sompolinsky, H., & Shapley, R. (1997). New perspectives on the mechanisms for orientation selectivity. *Current Opinion in Neurobiology, 7*, 514–522.

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

Thornber, K. K., & Williams, L. R. (2000). Characterizing the distribution of completion shapes with corners using a mixture of random processes. *Pattern Recognition, 33*, 543–553.

von der Heydt, R., Peterhans, E. & Baumgartner, G. (1984). Illusory contours and cortical neuron responses. *Science, 224*, 1260–1262.

Williams, L. R., & Jacobs, D. W. (1997a). Stochastic completion fields: A neural model of illusory contour shape and salience. *Neural Computation, 9*(4), 837–858,

Williams, L. R., & Jacobs, D. W. (1997b). Local parallel computation of stochastic completion fields. *Neural Computation, 9*(4), 859–881.

Zweck, J. W., & Williams, L. R. (2000). Euclidean group invariant computation of stochastic completion fields using shiftable-twistable functions. In *Proc. of the 6th European Conf. on Computer Vision (ECCV '00)*. Dublin, Ireland.

Received January 19, 1999; accepted October 3, 2000.