Stochastic Completion Fields: A Neural Model of Illusory Contour Shape and Salience

Lance R. Williams¹  David W. Jacobs
NEC Research Institute
4 Independence Way
Princeton, NJ 08540

Abstract

We describe an algorithm and representation level theory of illusory contour shape and salience. Unlike previous theories of this phenomenon, our model is derived from a single assumption—namely, that the prior probability distribution of boundary completion shape can be modeled by a random walk in a lattice whose points are positions and orientations in the image plane (i.e. the space which one can reasonably assume is represented by neurons of the mammalian visual cortex). Our model does not employ numerical relaxation or other explicit minimization, but instead relies on the fact that the probability that a particle following a random walk will pass through a given position and orientation on a path joining two boundary fragments can be computed directly as the product of two vector-field convolutions. We show that for the random walk we define, the maximum likelihood paths are curves of least energy, that is, on average, random walks follow paths commonly assumed to model the shape of illusory contours. A computer model is demonstrated on numerous illusory contour stimuli from the literature.

Keywords: Perceptual organization, grouping, figural completion, curve of least energy, occlusion, illusory contours, Markov process.

¹E-mail: williams@research.nj.nec.com
1 Introduction

The fundamental assumption underlying our work is that random walks in a lattice whose points are positions and orientations in the image plane can serve as models of the prior distributions of the shape of boundary completions. This is the space commonly assumed to be represented by neurons of the mammalian visual cortex. The completion shape and salience problem is the problem of computing the shape and relative likelihood (as determined by the prior distribution) of the family of curves which potentially connect (or complete) a set of contrast edges. This is a necessary intermediate step in the solution of the full figural completion problem, which has been previously characterized[Williams94c] as the problem of building a Huffman-labeled figure representing the boundaries of hidden and visible surfaces.

Although primarily envisioned as a cognitive theory, our work is also motivated by problems from computer vision, where perceptual organization is the key to efficient algorithms for solving a wide range of visual tasks. Many visual tasks become more efficient when computational effort can be focused selectively on those portions of an image likely to belong to a single object. For example, conventional approaches to object recognition search through a large number of combinations of image features, looking for subsets that match the features of a stored model. Lowe[Lowe85] pointed out that the combinatorial complexity of this search can be vastly reduced if one can determine, in advance, which subsets of image features are likely to belong to a single object. Jacobs[Jacobs89] has demonstrated that prior grouping can speed-up recognition by a factor of hundreds or thousands (see also [Burns93,Clemens91,Grimson90]). At the same time, many cognitive theories of human object recognition (e.g. [Biederman85]) also assume a good segmentation of the scene. If these theories are to plausibly model human performance, given the presence of occlusion and sensing noise, then methods of combining edges into cohesive wholes must be found. This is precisely the problem that we address. Apart from object recognition, perceptual organization has also been used to increase the efficiency of stereo correspondence (e.g. [Mohan89]) and motion segmentation and tracking (e.g. [Sawhney92]). Because our work addresses the problem of determining the likelihood that a set of fragmented contrast edges form the boundary of a single object, the output of our system is potentially suitable for all of these
purposes.

In the experimental implementation of the computational theory of figural completion described by Williams and Hanson[Williams94a,b] potential boundary completions are represented explicitly by tokens in a database. Completion shape is modeled by the cubic Bezier spline of least energy, and salience (or relative likelihood) was defined to be a function of this shape. Although it supported the computational theory, this implementation was not intended to be a realistic model at the level of algorithm and representation. In contrast, here we use a parallel distributed representation based upon a value unit encoding scheme, to represent potential boundary completions in a manner consistent with current knowledge of cortical structure. Like neurons of the mammalian visual cortex[Hubel62], the receptive field of each value unit is narrowly tuned to a specific position and orientation (i.e. a point in the space $R^2 \times S^1$). If we assume that the prior probability distribution of boundary completion shapes can be modeled by a random walk, then each value unit can be used to represent the probability that a Markov process is in a specific state. Let there exist two subsets of states, $P$ and $Q$, representing the beginning and ending points of a set of contrast edges corresponding to visible surface boundaries. We call $P$ the set of sources and $Q$ the set of sinks. Our goal is to compute the probability that a particle, initially in state $(x_p, y_p, \theta_p)$, for some $p \in P$, will in the course of a random walk (representing a prior on completion shape) pass through state $(u, v, \phi)$, on its way to state $(x_q, y_q, \theta_q)$, for some $q \in Q$, for all combinations of $u, v$ and $\phi$. We call this parallel distributed representation of the family of curves which potentially connect the set of contrast edges a stochastic completion field.\footnote{Although the stochastic completion field is based upon a Markov process, it is not a Markov Random Field (MRF).}

## 2 Previous Work

The relevant literature can be divided into three general categories: 1) computational theories of figural completion; 2) algorithm and representation level theories of shape and saliency; and 3) related methods from computer vision. The first provides the context necessary to understand our work as a cognitive theory; the latter two have similiar goals and scope. By computational theory of figural completion, we mean a detailed analysis of the computational
Figure 1: Three stages of figural completion (from [Williams94a]). Left: Input stimulus. Middle: boundary fragments (thick) and potential boundary completions (thin) represented by cubic Bezier splines of least energy. The stochastic completion field is an equivalent parallel distributed representation. Right: The labeled knot diagram is a viewer-centered representation of hidden and visible surfaces. The problem of deriving an equivalent parallel distributed representation is a subject for future research.

goal, natural constraints and inherent ambiguities, leading to a definition of the function which maps the available input to a representation sufficient to explain the observed human competence. Ideally, this function is described independently and in advance of an algorithm and representation level theory[Marr82].

Computational theories of figural completion are described in Kellman and Shipley[Kellman91] Nitzberg and Mumford[Nitzberg90], and Williams and Hanson[Williams94a,b]. These three theories are in many ways quite similiar. For example, all three exploit the fact that figural completion can be reduced to a purely combinatorial optimization problem by first committing to a set of potential completions of fixed plausible shape (see Figure 1). However, Williams is clearest in defining the computational goal, which he suggests is a viewer-centered representation of hidden and visible surfaces called a labeled knot diagram. A labeled knot diagram consists of one or more closed plane curves satisfying a subset of a labeling scheme first proposed by Huffman[Huffman71] for line drawings of smooth surfaces. Williams shows that the topological validity of the completed surfaces can be ensured by enforcing the labeling scheme, and that consequently, the labeling scheme represents an important source of grouping constraints.
To better appreciate the scope of the current paper, it will be useful to identify those phenomena which the stochastic completion field can account for, and those which it can not. For example, if one equates salience with apparent brightness (i.e. the perceived brightness of an illusory contour), then despite its title, this paper can not be a complete theory of salience. This is because apparent brightness depends on factors outside the scope of this theory, in particular, the topology of surfaces, of which our diffusion process has no knowledge. It is clear that apparent brightness is not solely a function of local configurational factors, (which we propose determine salience), but instead, is at least partly a function of the role the potential completion plays in the larger surface organization. Specifically, the apparent brightness is also a function of 1) whether or not the completion can be incorporated in a consistent way into a labeled knot diagram; and 2) whether the completion is nominally visible (i.e. modal) or occluded (i.e. amodal). Neither of these can be determined by a process which does not consider the topology of the scene.\(^3\)

Although the definition of stochastic completion field is quite specific, and explicitly embodies our assumption that the prior distribution of completion shape can be modeled as a random walk, other researchers have advanced theories of figural completion and image contour grouping based upon orientation fields of various kinds. Three of the earliest are due to Ullman[Ullman76], Grossberg and Mingolla[Grossberg87] and Parent and Zucker[Parent89]. At the level of algorithm and representation, all of these models (including the more recent models described below) are outwardly similar. This is because all derive from similar views of the orientation preference structure of the visual cortex, common assumptions about neural computation, and basic considerations (whether explicit or implicit) like translation and rotational invariance.

Like us, Heitger and von der Heydt[Heitger93] describe a theory of figural completion based upon non-linear combination of the convolutions of “keypoint” images with a fixed set of oriented grouping filters. Significantly, they demonstrate their method on both illusory contour figures like the Kanizsa triangle and on more “realistic” images (e.g. of plants and rocks) with impressive results. Unfortunately, neither the equations defining the filters nor

\(^3\)Consequently, the problem of deriving a parallel distributed equivalent of the labeled knot diagram lies outside the scope of this paper.
the proposed method of combination are described as a means of computing an underlying function. This makes analysis of their method difficult. However, it is worth noting that because their grouping filters are scalar functions, they can not (even implicitly) model a prior probability distribution of completion shapes in the manner we describe.

In a recent paper, Guy and Medioni[Guy93] describe a method for computing a vector-field representing global image structure from local tangent measurements. Like the Hough transform, the key to their approach is the local summation of a set of global voting patterns. Unlike the Hough transform, the accumulator is spatially registered with the image and the voting pattern is a vector-field, not a scalar-field. Elements of the vector-field represent orientations which are co-circular to the tangent measurements. The vectors are combined locally through analysis of moments, and the principal axis of the vectors which accumulate at a location is used as an estimate of the dominant direction. However, because this field is non-stochastic (i.e. deterministic), it can not model the prior distribution of completion shapes. Furthermore, their use of analysis of moments to combine votes of different orientations is an immediate consequence of their decision to use a non-stochastic field. That is, given that they desire a single orientation everywhere, and given that (in their scheme) evidence takes the form of votes of different orientations, in the end, they decide to compute an average.

While not portrayed as a cognitive theory, Shashua and Ullman[Shashua88] describe a network algorithm for computing “perceptual saliency.” Their method begins with an edge map, and iterates \( n \) times to produce a salience function over a set of discrete positions and orientations. Computing salience (as they define it) requires finding the energy of the minimum energy curve of length \( n \) beginning at every position and orientation. This energy function combines a term similar to squared curvature and a term measuring gap size. However, because the energy function also includes terms designed to guarantee convergence, the output of their system can be difficult to anticipate. Alter and Basri[Alter95] analyze the behavior of this network in detail, and point out some of its counterintuitive behavior. For example, they show that the saliency network may assign greater salience to a short texture line near a circle, than to the elements of the circle itself, and that it will assign greater salience to a circle with one large gap, than to a circle with several small gaps. In summary, although our work and Shashua and Ullman are in some ways similar, our work
differs because our choice of computational goal embodies a simple assumption about the stochastic nature of contour formation in the world.

## 3 Fundamental Assumption

Although our analysis is quite general, and remains valid for any translation and rotation invariant random walk in the space $R^2 \times S^1$, some random walks are likely to model the prior distribution of completion shapes better than others. In this paper, we define a random walk, which apart from satisfying the invariance criteria, on average generates short, smooth paths. For this random walk, it can be shown that the maximum likelihood path followed by a particle joining two points at fixed orientations is a curve of least energy (where energy is a linear combination of the integral of curvature squared and length). This is the curve which, others have theorized[Horn80,Nitzberg90,Ullman76], models the shape of illusory contours joining boundary fragments with orientation difference significantly less than $\frac{\pi}{2}$.

Unlike the familiar two-dimensional isotropic random walk, where a particle’s state is simply its position in the plane, the particles of the random walk introduced here possess both position and orientation. The random walk itself is defined by two elements: 1) the equations of motion; and 2) a decay constant. While the equations of motion describe the change in a particle’s position and orientation as a function of time, the decay constant, $\tau$, specifies a particle’s average lifetime. As part of a study of the reduction of search made possible by prior grouping in visual object recognition, Jacobs[Jacobs89] computed the probability density of the size of boundary gaps due to occlusion in random juxtapositions of a set of flat polygonal surfaces. Jacobs concluded that small gaps predominate and that incident frequency rapidly drops off with increasing size. By including a decay mechanism in the stochastic process, we are able to model the component of the completion shape probability distribution dependent on length. Because a certain fraction of particles (i.e. $1 - e^{-\frac{t}{\tau}}$) decay per unit time, longer paths are exponentially less likely. A particle’s position and orientation are updated according to the following stochastic non-linear differential equation:

\[
\dot{x} = \cos \theta
\]
\[
\begin{align*}
\dot{x} &= \sin \theta \\
\dot{\theta} &= \hat{\kappa}(0, \sigma^2; t)
\end{align*}
\]

where \( \dot{x} \) and \( \dot{y} \) specify change in position, \( \dot{\theta} \) is change in orientation (i.e. curvature) and \( \hat{\kappa}(0, \sigma^2; t) \) is a normally distributed random variable with zero mean and variance equal to \( \sigma^2 \). There is a strong similarity between the equations of motion defined here and the Kalman filter equations Cox, Rehg and Hingorani[93] employed in their work on edge tracking—a particle moves with constant speed in a direction which is continually changing by some random amount. The effect is that particles tend to travel in straight lines, but over time, drift to the left or right by an amount proportional to \( \sigma^2 \) (for example random walks, see Figure 2). When \( \sigma^2 = 0 \), the motion is completely deterministic, and particles never deviate from straight paths. When \( \sigma^2 = \infty \), the motion is completely random, and the stochastic process becomes a two-dimensional isotropic random walk. For this reason, and although these equations of motion differ from those conventionally used to model the motion of a particle of gas in a vacuum, the analogy is useful, and we will sometimes refer to the stochastic process we define as a diffusion process, and the parameter, \( \sigma^2 \), as the diffusivity.
In this section, we demonstrate the relationship between the diffusion process that we have described, and so-called curves of least energy[Bruckstein90,Horn81,Ullman76]. To accomplish this, we employ a discrete-time approximation to the continuous equations of motion. For the discrete-time random walk, we show that the most likely path between a source and a sink is the curve of least energy connecting them, for a natural, discrete formulation of energy. This tells us that, in the limit, as the time-step size becomes small, the most likely path between a source and a sink is the continuous curve of least energy. From this, we conclude that our model of diffusion concisely encodes the prior assumption that the most likely shape of an occluded section of an object’s boundary is the curve of least energy.

Suppose that a particle begins its random walk at a source, \( p \), and follows a trajectory \( \Gamma \), which consists of \( n \) unit length steps, along with \( n \) changes in angle denoted by \( \kappa_1, \ldots, \kappa_n \). That is, its trajectory is an \( n \)-sided polygonal arc, comprised of unit length segments and with exterior angles denoted by \( \kappa_i \). From the definition of the random walk, we know that the density function on the set of paths that such a particle may take is given by:

\[
f(\Gamma_p) = \prod_{i=1}^{n} e^{-\frac{1}{\sigma^2} \frac{k_i^2}{2\sigma^2}}
\]

Let \( g(\Gamma_{pq}) \) denote a density function on the set of possible paths which begin at source \( p \) and end in sink \( q \). Then we have:

\[
g(\Gamma_{pq}) = \frac{1}{\int_{\Gamma_q} f(\Gamma_p) d\Gamma_p} \prod_{i=1}^{n} e^{-\frac{1}{\sigma^2} \frac{k_i^2}{2\sigma^2}}
\]

where \( \int_{\Gamma_q} f(\Gamma_p) d\Gamma_p \) indicates integration of the density function \( f(\Gamma_p) \) (i.e. all paths beginning in source \( p \)) over all paths ending in sink \( q \) (see Ash[Ash72], or a similar text for a discussion of conditional probabilities on continuous functions). Taking the logarithm of both sides we get:

\[
\log(g(\Gamma_{pq})) = -\log(\int_{\Gamma_q} f(\Gamma_p) d\Gamma_p) + n(-\frac{1}{\tau} - \log(\sigma \sqrt{2\pi})) - \sum_{i=1}^{n} \frac{k_i^2}{2\sigma^2}
\]

which may be rewritten as:

\[
\log(g(\Gamma_{pq})) + \log(\int_{\Gamma_q} f(\Gamma_p) d\Gamma_p) = n(-\frac{1}{\tau} - \log(\sigma \sqrt{2\pi})) - \sum_{i=1}^{n} \frac{k_i^2}{2\sigma^2} \tag{1}
\]
or because the second term of (1) is constant for given \( p \) and \( q \):

\[
\log(g(\Gamma_{pq})) + C = n(-\frac{1}{\tau} - \log(\sqrt{2\pi})) - \sum_{i=1}^{n} \frac{k_i^2}{2\sigma^2} \tag{2}
\]

We now relate this expression to the energy of a continuous curve. This energy is defined as:

\[
\alpha \int_{\Gamma} \kappa(t)^2 dt + \beta \int_{\Gamma} dt
\]

where \( \kappa(t) \) is the curvature at \( \Gamma(t) \) and \( \alpha \) and \( \beta \) are constants that weight the cost for the length relative to the cost for the total squared curvature. Although the energy of any curve with a curvature discontinuity is infinite, it is natural to consider an analogous quantity for polygonal arcs based on 1) number of segments (i.e. \( n \)); and 2) sum of the squared exterior angles (i.e. \( \sum_{i=1}^{n} k_i^2 \)). This is precisely the negative of the right side of equation (2), with \( \alpha = 1/2\sigma^2 \) and \( \beta = 1/\tau + \log(\sqrt{2\pi}) \). This shows that the log-likelihood that a random walk will follow any given polygonal path is linearly related to the energy of that path for a natural discrete formulation of energy. Therefore, the maximum likelihood polygonal path between two points is the discrete curve of least energy.

It is worth noting that while the discrete-time random walk analyzed here is computationally expedient (see section 5), it is not mathematically necessary. Furthermore, the two are equivalent in the limit, because the discrete-time random walk becomes a continuous-time random walk as the time-step size, velocity and diffusivity approach zero. In this case, the maximum likelihood path is the continuous curve of least energy. It follows that the continuous equations of motion concisely encode our prior assumption concerning the probability distribution of completion shape.

4 Continuous Problem Formulation

The receptive fields of neurons in the visual cortex are spatially localized and narrowly tuned to stimuli with particular orientations[Hubel62]. It follows that the activity of neurons of the visual cortex can be represented by a probability density function over the three-dimensional space, \( R^2 \times S^1 \). Furthermore, the network of interconnections among the neurons of the cortex can be represented by a rank six tensor, \((R^2 \times S^1) \times (R^2 \times S^1)\). Conservatively speaking, the function the network computes can be viewed as a tensor product, mapping
an input probability density function to an output probability density function through a linear transformation.

Let the rank six tensor $G$ represent the transition probabilities of a Markov process defined on the three-dimensional state space, $R^2 \times S^1$, and satisfying the equations of motion defined above (i.e. $G$ is the Green's function). It follows that $G(u, v, \phi, x, y, \theta ; t)$ is the probability that a random walk of length $t$ will end in state $(u, v, \phi)$ given that it began in state $(x, y, \theta)$. If $p(x, y, \theta ; 0)$ is the probability density function describing the position and orientation of a particle at the beginning of its random walk, then the probability that the particle will be at some position and orientation at time $t$ is:

$$p(u, v, \phi ; t) = \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy \int_{-\pi}^{\pi} d\theta \ G(u, v, \phi, x, y, \theta ; t) p(x, y, \theta ; 0) \cdot e^{\frac{-t}{2}}$$

Because the probability of a random walk of a given shape is independent of its initial position and orientation in the plane, the rank six tensor, $G$, consists entirely of translated and rotated copies of a rank three tensor representing the transition probabilities for random walks beginning at $(0, 0, 0)$:

$$G(u, v, \phi, x, y, \theta ; t) = G(u' - x, v' - y, \phi - \theta, x - x, y - y, \theta - \theta ; t)$$

$$= G(u' - x, v' - y, \phi - \theta 0,0,0; t)$$

where $(u', v', \phi - \theta)$ is $(u, v, \phi)$ rotated by $-\theta$ about $(x, y)$, so that $u' = u \cos \theta + v \sin \theta$ and $v' = -u \sin \theta + v \cos \theta$. Henceforward, we will write $G(u' - x, v' - y, \phi - \theta ; t)$ instead of $G(u' - x, v' - y, \phi - \theta, 0,0,0 ; t)$. The upshot is, that by taking advantage of the translational and rotational symmetries of $G$, the tensor product can be computed by convolution:

$$p(u, v, \phi ; t) = \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy \int_{-\pi}^{\pi} d\theta \ G(u' - x, v' - y, \phi - \theta ; t) p(x, y, \theta ; 0) \cdot e^{\frac{-t}{2}}$$

Critical to our result, is the fact that the stochastic completion field can be expressed as the product of a stochastic source field and a stochastic sink field. We define the stochastic source field, $p'(u, v, \phi)$, to be the fraction of paths which begin in a source state and pass
Figure 3: Upper left: P.d.f. representing source distribution, \( p(x, y, \theta ; 0) \), consists of an impulse located at \((0, 0, 0)\). Although sometimes truncated for clarity, in general, the length of the arrows is proportional to the logarithm of the probability that a particle is located at that position and orientation. Upper right: Snapshot of diffusion process at \( t = 15 \), that is, \( p(x, y, \theta ; 15) \). This is the result of convolving \( G(u', v', \phi ; 15) \) with the p.d.f. representing \( t = 0 \). Lower left/right: Diffusion process at \( t = 31 \) and \( t = 47 \), respectively.
through \((u, v, \phi)\) before they decay. Note that the source field differs from the completion field because the path is not required to end in a sink state. If we assume that paths do not self-intersect before they decay,\(^4\) then the fraction of paths which pass through a given position and orientation equals the integral of the position and orientation probability density function over time:

\[
p'(u, v, \phi) = \int_0^\infty dt \ p(u, v, \phi ; t) \\
= \int_0^\infty dt \int_{-\infty}^\infty dx \int_{-\infty}^\infty dy \int_0^\pi d\theta \ G(u' - x, v' - y, \phi - \theta ; t) \ p(x, y, \theta ; 0) \cdot e^{-\frac{t}{\tau}}
\]

By changing the order of integration, and defining a new Green’s function, \(G'\):

\[
G'(u, v, \phi) = \int_0^\infty dt \ G(u, v, \phi ; t) \cdot e^{-\frac{t}{\tau}}
\]

the explicit time variable, \(t\), can be suppressed. The result is that the stochastic source field, \(p'(u, v, \phi)\), can be computed by convolving the p.d.f. representing the source distribution with the new Green’s function:

\[
p'(u, v, \phi) = \int_{-\infty}^\infty dx \int_{-\infty}^\infty dy \int_0^\pi d\theta \ G'(u' - x, v' - y, \phi - \theta) \ p(x, y, \theta ; 0)
\]

With the expression for \(p'(u, v, \phi)\) in hand, we are now in a position to formulate an expression for the completion field, \(C(u, v, \phi)\). Observe that the probability that a particle will pass through state \((u, v, \phi)\) on its way from a source state, \(p \in P\), to a sink state, \(q \in Q\) before it decays, is proportional \(^5\) to the product of 1) the probability that a particle

\(^4\)Actually, we need only worry about particles which return to a state they have already visited, that is, paths which cross themselves with matching orientations. Because the probability of the second visit is conditional on the first visit, these would be “double-counted” by the integration. Fortunately, it is easy to show (experimentally) that for a wide range of diffusivities and decay constants the chance of this happening is quite small.

\(^5\)Strictly speaking, what we are able to compute is a relative likelihood, not a probability. This is because we do not compute the absolute probability of a particle diffusing from a source to a sink by any path at all, which would represent the normalizing constant. So, although it is possible to convert these relative likelihoods to probabilities, in practice, this is not necessary, because relative likelihoods suffice for the purpose of comparing competing paths.
Figure 4: Stochastic source fields, $p'(u, v, \phi)$, representing the probability that a particle leaving $(0,0,0)$ will reach $(u, v, \phi)$ before it decays. Basically, these are plots of the Green’s function $G'$ (i.e. the time integral of $G$), for a range of diffusivities and decay constants. Upper left: $(\sigma^2 = 0.05, \tau = 100)$. Upper right: $(\sigma^2 = 0.05, \tau = 20)$. Lower left: $(\sigma^2 = 0.01, \tau = 100)$. Lower right: $(\sigma^2 = 0.01, \tau = 20)$. 
beginning in a source will reach \((u, v, \phi)\) before it decays (i.e. the source field); and 2) the probability that a particle beginning at \((u, v, \phi)\) will reach a sink before it decays (i.e. the sink field). We have shown that the source field, \(p'(u, v, \phi)\), can be computed by convolving the source distribution with the Green’s function, \(G'\). We now show that the sink field can be computed in a similar way.

Observe that the probability that a particle undergoing a random walk will follow some path is independent of the direction in which the path is traversed. More formally, the probability that a particle leaving \((x, y, \theta)\) will reach \((u, v, \phi)\) before it decays is equal to the probability that a particle leaving \((u, v, \phi + \pi)\) will reach \((x, y, \theta + \pi)\) before it decays:

\[
G'(u, v, \phi, x, y, \theta) = G'(x, y, \theta + \pi, u, v, \phi + \pi)
\]

Adopting the convention that \(\overline{q}'(u, v, \phi)\) represents the probability that a particle leaving \((x_q, y_q, \theta_q + \pi)\), for \(q \in Q\), will reach \((u, v, \phi)\) before it decays, then:
\[
\bar{q}'(u, v, \phi) = \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy \int_{-\pi}^{\pi} d\theta \ G'(u' - x, v' - y, \phi - \theta) \ q(x, y, \theta + \pi ; 0)
\]

The sink field proper, \( \bar{q}'(u, v, \phi + \pi) \), which represents the probability that a particle leaving \((u, v, \phi)\) will reach a sink state before it decays, is then computed by substituting \( \phi + \pi \) for \( \phi \) in the above equation:

\[
\bar{q}'(u, v, \phi + \pi) = \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy \int_{-\pi}^{\pi} d\theta \ G'(u' - x, v' - y, \phi + \pi - \theta) \ q(x, y, \theta + \pi ; 0)
\]

Finally, the stochastic completion field, \( C(u, v, \phi) \), which represents the relative likelihood that a particle leaving a source state will pass through \((u, v, \phi)\) and enter a sink state before it decays, equals the product of the source and sink fields:

\[
C(u, v, \phi) = p'(u, v, \phi) \cdot \bar{q}'(u, v, \phi + \pi)
\]

\[
= \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy \int_{-\pi}^{\pi} d\theta \ G'(u' - x, v' - y, \phi - \theta) \ p(x, y, \theta ; 0)
\]

\[
\cdot \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy \int_{-\pi}^{\pi} d\theta \ G'(u' - x, v' - y, \phi + \pi - \theta) \ q(x, y, \theta + \pi ; 0)
\]

5 Discrete Problem Formulation

Although the continuous formulation is of theoretical interest, a discrete formulation is a more realistic model of the visual cortex and has the further advantage of simplifying the computer implementation. There are two ways in which the discrete formulation can be achieved. The first is to redefine the continuous-state, continuous-time Markov process used to model the completion shape prior as a discrete-state, discrete-time Markov chain. However, the advantage of the continuous formulation is the simplicity of the equations of motion (which embody the prior assumption), and it seems unlikely that one could define a suitable prior as transition probabilities on a discrete lattice in as straightforward a way.\(^6\)

We therefore employ a discrete approximation to the solution of the continuous equations

\(^6\)However, it is quite plausible that a prior distribution in the form of transition probabilities could be learned from a large set of example images.
Figure 6: Upper left: The stochastic source field, $p'(u, v, \phi)$, represents the probability that a particle will leave a source state and reach state $(u, v, \phi)$ before it decays. Upper right: The stochastic sink field, $\bar{q}'(u, v, \phi + \pi)$, represents the probability that a particle will leave state $(u, v, \phi)$ and enter a sink state before it decays. Bottom: The stochastic completion field, $C(u, v, \phi)$, is the product of source and sink fields.
of motion. Although it is possible to solve the stochastic non-linear differential equation directly, and in doing so, find an analytic equation for the Green’s function (Thornber and Williams[Thornber95]), it is considerably simpler to solve for \( G' \) by a Monte Carlo method. All of the experimental results in this paper are based upon an approximation of \( G' \) computed by direct simulation of the random walk for \( 1.0 \times 10^6 \) trials on a \( 256 \times 256 \) grid with 36 fixed orientations. The probability that a particle beginning at \((0,0,0)\) will reach \((u,v,\phi)\) before it decays (i.e. \( G'(u,v,\theta) \)) is approximated by the fraction of simulated trajectories beginning at \((0,0,0)\) which intersect the region \((u \pm 1.0, v \pm 1.0, \phi \pm \pi/72)\).

Convolution of a source (or sink) distribution of size \( n \times n \times m \) with a kernel of the same size (i.e. representing \( G' \)) requires \( O(n^3m^2) \) operations on a sequential computer. However, if the source and sink distributions consist of a relatively small number of impulses, then it is more efficient to synthesize the source and sink fields from the sum of translated and rotated copies of the convolution kernel. This is the key to the efficiency of the voting pattern method proposed by Guy and Medioni[Guy93]. For a source or sink distribution consisting of \( O(k) \) impulses, this requires only \( O(n^2mk) \) operations, yielding significant computational savings.

6 Experimental Results

As a first demonstration, let us consider three source/sink pairs. The source and sink distributions were represented by arrays of size \( 128 \times 128 \times 36 \) and consisted of a single oriented impulse (i.e. a single non-zero value). In each case, source and sink are positioned on a horizontal axis and are oriented symmetrically about this axis. In all the experiments in this paper, the convolution kernel representing \( G' \), was of size \( 256 \times 256 \times 36 \). The diffusivity, \( \sigma^2 \), equaled 0.05 and the decay constant, \( \tau \), equaled 20. The locations of the oriented impulses comprising the first source/sink pair are \( p_1 = (-40, 0, 10^\circ) \) and \( q_1 = (40, 0, -10^\circ) \). The locations of those comprising the second are \( p_2 = (-40, 0, 30^\circ) \) and \( q_2 = (40, 0, -30^\circ) \) and the third are \( p_3 = (-40, 0, 50^\circ) \) and \( q_3 = (40, 0, -50^\circ) \). Figure 7 depicts the stochastic completion field for these three source/sink pairs as brightness images where brightness encodes the sum over all 36 orientations. Because, for each pair, the displayed brightnesses
are scaled to take maximum advantage of the limited number of grey levels (i.e. 255), it is not obvious that there is a two order of magnitude difference in saliency between the first and last pair. The first source/sink is scaled by $1.0 \times 10^6$, the second by $1.0 \times 10^7$ and the third by $1.0 \times 10^8$. The scale factors approximately equal the multiplicative inverse of the maximum likelihood.

As a second demonstration, we again use three source/sink pairs. In each case, source and sink are positioned on a horizontal axis and possess the same orientation, so that the curves of least energy joining them will contain inflection points. The locations of the oriented impulses comprising the first source/sink pair are $p_1 = (-40,0,10^\circ)$ and $q_1 = (40,0,10^\circ)$. The locations of those comprising the second are $p_2 = (-40,0,30^\circ)$ and $q_2 = (40,0,30^\circ)$ and the third are $p_3 = (-40,0,40^\circ)$ and $q_3 = (40,0,40^\circ)$. Figure 8 depicts the stochastic completion field for these three source/sink pairs as brightness images where brightness encodes the sum over all 36 orientations. Again, the displayed brightnesses are scaled to take maximum advantage of the limited number of grey levels. The first source/sink is scaled by $1.0 \times 10^7$, the second by $1.0 \times 10^7$ and the third by $1.0 \times 10^8$.

In a third demonstration, we consider a source distribution consisting of four oriented impulses equally spaced around the circumference of a circle (see Figure 5 (left)). This distribution is meant to represent an Ehrenstein figure (see Figure 9 (left)). The four impulses are located at endpoints of the four line segments comprising the figure and possess orientation normal to the segments. Figure 9 (right) shows the stochastic completion field, where brightness encodes the sum over all orientations.

We have also demonstrated our implementation on several well known illusory contour figures from the visual psychology literature. Before we could do this, we had to figure out a principled way of translating an image into a set of sources and sinks for our diffusion process. In general, this requires a sophisticated analysis of the local image brightness structure to identify “L-junctions,” “T-junctions,” “Y-junctions,” and “X-junctions” formed by both contrast and outline edges. Classifying and measuring the multiple orientations at so called “keypoints” is a difficult research problem in its own right and the subject of much current research[Freeman91, Heitger93, Michaelis94, Perona92]. For the moment, we use a steerable one-sided filter scheme similar to that of Michaelis[Michaelis94] to identify
Figure 7: Example completions. Left: $p_1 = (-40, 0, 10^\circ)$ and $q_1 = (40, 0, -10^\circ)$. Middle: $p_2 = (-40, 0, 30^\circ)$ and $q_2 = (40, 0, -30^\circ)$. Right: $p_3 = (-40, 0, 50^\circ)$ and $q_3 = (40, 0, -50^\circ)$.

orientation discontinuities (i.e. corners) formed by contrast edges and to measure the two orientations with precision sufficient for our purposes. The maximum and minimum of the continuous response of the steerable one-sided filter is first measured to approximately $3^\circ$ of accuracy. As an anti-aliasing measure, a unit mass is distributed proportionally among the two discrete orientations straddling the nominal maximum and minimum. The maximum is interpreted as a source and the minimum as a sink (see Figure 10 (top left and right)). Consequently, illusory contours joining contrast edges of opposite sign do not occur. Generalizing this scheme to classify and measure the wide range of events corresponding to likely points of boundary occlusion is the only obstacle to demonstrating our work on more “realistic” images.

Figure 10 (bottom left) shows the Kanizsa triangle stimulus. Keypoints are located at a positive maximum of curvature. Figure 10 (bottom right) shows the stochastic completion field summed over all orientations and superimposed on the brightness gradient magnitude image. Both the illusory triangle and the three discs are completed. In contrast with the results of Heitger and von der Heydt[Heitger93], no non-maximum suppression needs to be employed.
Figure 8: Completions with inflections. Left: $p_1 = (-40, 0, 10^\circ)$ and $q_1 = (40, 0, 10^\circ)$. Middle: $p_2 = (-40, 0, 20^\circ)$ and $q_2 = (40, 0, 20^\circ)$. Right: $p_3 = (-40, 0, 30^\circ)$ and $q_3 = (40, 0, 30^\circ)$.

Figure 11 (top) shows the Kanizsa “paisley” stimulus, which is a well known example of the Petter effect[Petter56]. The Petter effect occurs when two surfaces of equal reflectance overlap. Because the reflectances of the two surfaces are the same, their relative depth can not be determined from figural information alone. Even so, there is a strong tendency to see the broader of the two surfaces in front of the narrower (i.e. the longer completion is perceived amodally). Figure 11 (bottom) shows the stochastic completion field integrated over all orientations and superimposed on the brightness gradient magnitude image. Keypoints are located at negative minima of curvature. It is interesting to note that the average likelihoods of the shorter completions (i.e. perceived modally) are several orders of magnitude greater than the average likelihoods of the longer completions (i.e. perceived amodally).

Figure 12 (top) shows a complex illusory contour figure designed by Kanizsa[Kanizsa79]. This figure illustrates that whether or not an illusory contour is perceived is not solely a function of local configurational factors, but also depends on whether or not the completion can be incorporated in a globally consistent way in a labeled knot diagram. Because our diffusion process has no knowledge of the topology of surfaces, the stochastic completion field, shown in Figure 12 (bottom left), contains potential completions which are not perceived by
human subjects. Potential completions required to complete the four rectangles are among the most salient, however. Figure 12 (bottom right) shows the logarithm of the stochastic completion field integrated over all orientations. In the logarithm image, many additional completions of significantly lower average likelihood become visible. Included among these are those required to complete the four black discs and eight black squares perceived by human subjects.

7 Discussion

Although the contribution of this work can be described in different ways, one way which is worth mentioning is as a means of computing the curve of least energy interpolating a set of image measurements using a feedforward neural network (i.e. a network with no recurrent connections). Marr[Marr82] pointed out that neurons are slow and grouping is very fast, and that this seems to preclude the use of relaxation algorithms, which often converge quite slowly. Yet others have suggested that the shape of illusory contours is described by the curve of least energy[Horn81,Nitzberg90,Ullman76], which is normally computed by variational methods. So, how can an optimal shape emerge from a computation where
Figure 10: Top left: A corner of a “pacman.” Top right: Sources (thin) and sinks (thick) computed by steerable one-sided filter (multiple orientations are due to anti-aliasing). Bottom left: Kanizsa triangle. Bottom right: Stochastic completion field summed over all orientations and superimposed on brightness gradient magnitude image. Both the illusory triangle and the three discs are completed.
Figure 11: Top: Kanizsa’s “paisleys.” Bottom: Stochastic completion field integrated over all orientations and superimposed on brightness gradient magnitude image. It is interesting to note that the average likelihoods of the shorter completions (i.e., perceived modally) are several orders of magnitude greater than the average likelihoods of the longer completions (i.e., perceived amodally).
Figure 12: Top: A complex figure[1]. Bottom: Stochastic completion field integrated over all orientations and superimposed on brightness gradient magnitude image. Logarithm of stochastic completion field integrated over all orientations.
nothing is explicitly minimized? We suggest an analogy with Hamiltonian formulations in physics, for example, with the equations of motion of a particle in a gravitational field and the principle of least action. Nowhere in the equations of motion does a quantity representing action\(^7\) appear, yet the action is minimized by the path a particle takes. Similarly, nowhere in the stochastic non-linear differential equation governing the motion of a particle in our scheme does energy explicitly appear, yet we have shown that the maximum likelihood paths joining image measurements are curves of least energy.

8 Conclusion

It is widely acknowledged that perceptual organization (i.e. segmentation, grouping) is among the most difficult problems facing researchers in computer vision today. In our opinion, research in bottom-up visual reconstruction has virtually come to a standstill because of lack of progress in this area. The development of practical object recognition systems has also been seriously hampered. Our work advances the state of the art in perceptual organization by providing a plausible model of illusory contour formation based upon a diffusion process. We have shown how curves of least energy interpolating a set of edge fragments can be computed using biologically plausible algorithms and representations. Significantly, this is accomplished without using numerical relaxation or other explicit minimization, but instead relies on the fact that the probability that a particle following a random walk will pass through a particular position and orientation on a path joining two image measurements can be computed directly as the product of two vector-field convolutions.

The most related work is by Heitger and von der Heydt[Heitger93], Guy and Medioni[Guy93] and Shashua and Ullman[Shashua88]. Although we owe an intellectual debt to these three papers, none of these papers proceeds cleanly from a definition of what they want to compute (i.e. a function) to a method of computing it (i.e. an algorithm). Our main contribution is to show how outwardly similar algorithms and representations follow immediately from a clearly stated assumption—namely, that the prior probability distribution of boundary completion shapes can be modeled by a random walk.

\(^7\)The difference between potential and kinetic energies[Feynmann64].
Acknowledgments The authors gratefully acknowledge many helpful conversations with Karvel Thornber. Thanks also to Ingemar Cox and Bill Bialek.

References


Burns93 Burns, J.B., Matching 2D Images to Multiple 3D Objects Using View Description Networks, Ph.D. Dissertation, Dept. of Computer Science, Univ. of Massachusetts at Amherst, Amherst, Mass., 1993.


Thornber95 Thornber, K.K. and L.R. Williams, Analytical Stochastic Completion Fields, forthcoming.


