In *Workshop on Statistical and Computational Theories of Vision*
at *International Conference on Computer Vision*, Vancouver, July 13, 2001.

A Generative Model of Curve Images with a Completely-Characterized Non-Gaussian Joint Distribution

Jonas August
jonas.august@yale.edu

Steven W. Zucker
steven.zucker@yale.edu

Center for Computational Vision and Control
Departments of Electrical Engineering and Computer Science
Yale University
51 Prospect St., New Haven, CT 06520

Abstract

What is an ideal edge map? Can one construct a probabilistic, generative model of images of contours that is tractable? Motivated by these questions, we define a prior model for ideal edge maps by assuming that they are generated by Markov processes via an indicator function. In this theoretical paper we analyze this *curve indicator random field* model both in the single curve and multiple curve cases. In particular, we derive exact, usable expressions for this generative model's moment generating functional as well as all of its joint cumulants. We show that this prior is non-Gaussian, and we outline how it can be combined with an observation model. The resulting filter requires the solution of two partial differential equations.

1 Introduction

Probabilistic models in vision fall roughly into two camps: one-dimensional and two-dimensional. Contour models supporting curve detection fall into the first class, while texture and segmentation models fall into the second. Someone seeking *generative* models for vision immediately realizes that these two classes are not equivalent: images are two-dimensional, and therefore require two-dimensional generative models. In this paper we study a generative model for curve images that explicitly provides a link between these two dimensionalities.

Previous work with one-dimensional models in vision was applied to contour detection. The earliest work was by Montanari[19], where the objective was to find a single contour in noise. To exploit the sequential structure of the problem, dynamic programming was used to solve for the optimal curve; later, Martelli studied similar problems with heuristic search [18]. Although these early models were not explicitly probabilistic, they motivated recent work on the tracking of roads in satellite images [12] that was. The Bayesian framework provided tools that enabled Yuille and Couglan to perform an asymptotic performance analysis [25]. Cox et al constructed a contour tracking algorithm that combined Kalman filtering with the multiple hypothesis generation of heuristic search [7]. Unfortunately, the latter model requires low noise, while the other models depend upon user input of a contour endpoint and the contour length, and work only for a single contour. None of these works provides a generative model of contour images.

The generative nature of two-dimensional probabilistic models in vision has supported a wide variety of tasks. Gaussian random fields have been applied to image restoration [15], to texture modeling [6], and, via principle components analysis, to face recognition [23]. Markov random fields have also been applied to texture modeling [8], segmentation [13], image restoration [5], and even contour modeling [17]. Difficulty with parameter estimation and phase transitions, coupled with observations of sample images of these field models motivated a more empirical approach, such as the Minimax Entropy framework for capturing the observed statistics of a variety of filter outputs from natural images [26]. The key blessing of these models is also their difficulty. Being very general they offer the promise of a comprehensive probabilistic account of images; however, computational complexity increases

dramatically when that generality is actually exploited, e.g., as neighborhood sizes increase in Markov random fields or unless joint filter statistics are ignored in the Minimax Entropy framework. Nonetheless, compelling images can be generated from these models, particularly of textures.

Here we focus on contour modeling, and so we in effect study only one class of filters: (ideal) local edge or line operators at positions (x, y) and directions θ. This simplification allows us to explicitly consider their interactions; in particular, we consider all the joint statistics, of all orders k, of a random field of ideal edges in (x, y, θ). Evolution seems to have exploited such interactions: witness the surge of activity in studying "horizontal" connections among simple cells in the primary visual cortex of primates and contextual effects beyond the classical receptive field [16]. More recently, we have measured edges correlations in natural images [2], demonstrating the importance of joint filter statistics.

Our object of study is an ideal field of local edge/line operator responses, a field generated by a random number of independent, randomly long, smooth curves. The novelty in this conception is that we map the set of contours (each merely a list of points) to a field, and to do this we employ an indicator function (Def. 1). The resulting (random) field is then directly comparable to the local edge/line operator responses. This construction is related to the line processes often used to separate regions in Markov random fields [13], although we do not pursue that connection here. We now consider our modeling assumptions in greater detail.

Markov processes are ideally suited to capture the local property of smoothness: for example, Mumford, and later Williams and co-workers, imagined a particle at $R_t = (X_t, Y_t, \Theta_t) \in \mathbb{R}^2 \times \mathbb{S} = \{(x, y, \theta)\}$ whose direction Θ_t is slightly perturbed at each time instant t before taking its next step forward. Mathematically, Mumford's Markov process has the stochastic differential equation:

$$\frac{dX}{dt} = \sin\Theta, \quad \frac{dY}{dt} = \cos\Theta, \quad d\Theta = \sigma dW, \tag{1}$$

where σ bounds the direction perturbations and W is standard Brownian motion (on the circle $\mathbb{S}$). In our framework the Markov process R_t models all image contours, some observed without corruption of any sort (e.g., no noise nor blur), some poorly observed (e.g., medical

images), and some invisible (e.g., occluded contours, Mumford's original application). Indeed, the particular (stationary) Markov process contour model is unspecified in our model; more exotic processes, which include scale [22] or curvature κ [4], can be used as well. At this level of generality, the Markov process R_t takes on values ("states") i in state space $\mathcal{I}$, e.g. $i = (x, y, \theta)$ or $i = (x, y, \theta, \kappa)$.

Several independence assumptions play a role in our model. First, as an approximation we assert a memoryless property: whether the contour continues beyond a certain point does not depend on how long it is already. This implies an exponential distribution over contour length T. Second, we observe that independence of contours is a reasonable approximation for modeling elongated curves. Tree branch contours, for example, statistically interact primarily where they meet; due to the local process of growth, the individual branches wander largely independently. Finally, the random number of contours is assumed to be Poisson-distributed: in the related context of random point fields, this distribution is natural where disjoint regions contain independent numbers of points. That such assumptions are reasonable can be seen in Fig. 1.

A real benefit of creating generative models of image abstractions is that the simplifications hold the promise of yielding to analysis. By focusing our attention on contour images, we obtain results much stronger than are possible for more general probabilistic frameworks. Therefore, this is a theoretical paper, where we:

1. Define an exact model for random contour images—the "curve indicator random field"—that is general enough to capture many notions of contour smoothness including orientation and curvature good continuation;

2. Analytically compute its joint cumulants of all orders, and derive its moment generating functional;

3. Provide a checkable convergence condition on the moment generating functional;

4. Show that the curve indicator random field is non-Gaussian and yet has a tractable form; and

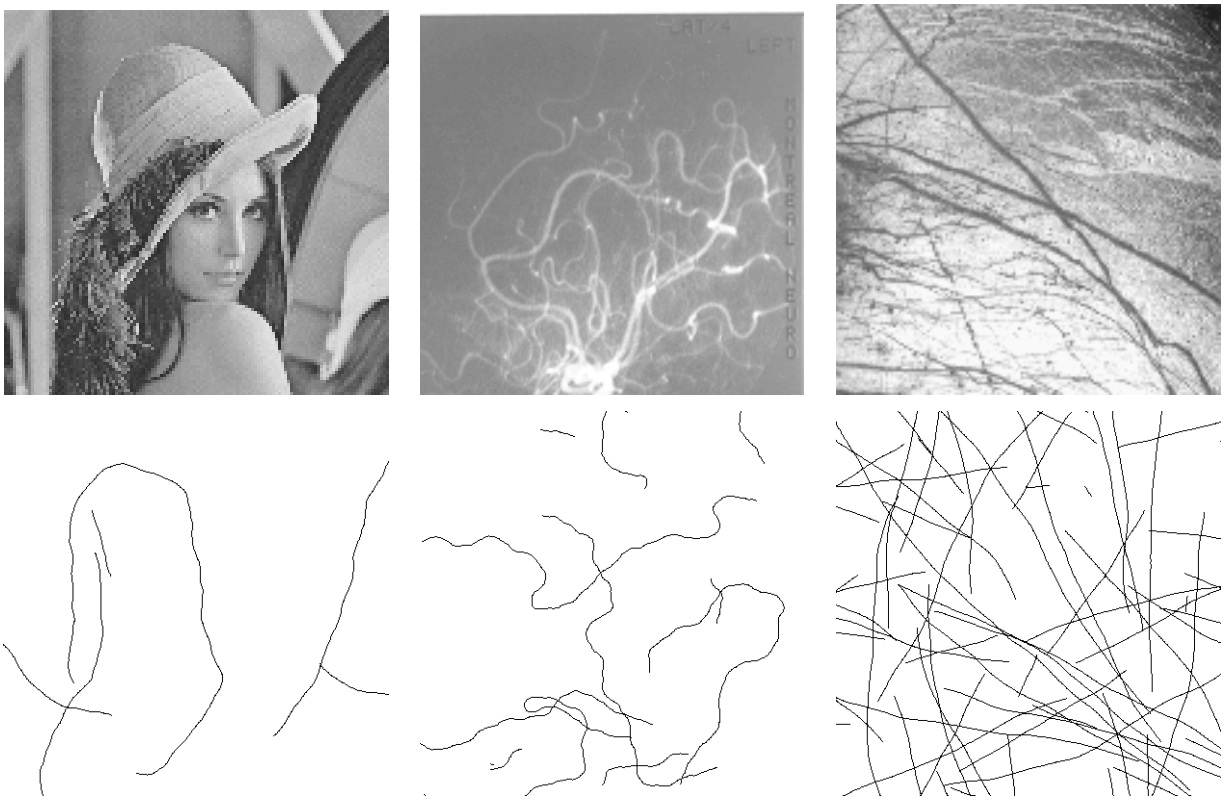

Figure 1: Observe the similarity of contours in natural images (top: "Lenna," angiogram, possible ice cracks on Jupiter moon Europa) to samples generated from the curve indicator random field (CIRF) for Mumford's Markov process [20] in (x, y, θ) (bottom: various parameter settings of CIRF). Although these images (top) arise from distinct physical sources, they all have contours, differing primarily in number, smoothness, and extent. The CIRF acts as a prior for contour enhancement in our framework.

5. Show how to combine this curve indicator random field prior with a noise model to define a *nonlinear* filter for contour enhancement.

The third result above is special in that it not only characterizes the domain of definition of the model, but provides a signal processing-like "stability" condition that we can readily verify. Accessible proofs are provided for all propositions.

2 The Curve Indicator Random Field

Given a Markov process for modeling individual contours as described in §1, we now define a *curve indicator random field* (CIRF), which naturally captures the notion of an ideal edge/line map. Roughly, this random field is non-zero-valued along the true contours, and zero-valued elsewhere. The actually measured edge/line map is then viewed as an *imperfect* CIRF, corrupted by noise, blur, etc. Because the CIRF is not standard, our presentation will be self-contained.

2.1 Definitions

For generality, we shall define the curve indicator random field for any stationary continuous-time Markov process $R_t, 0 \leq t < T$ taking values in a finite (or at most countable) set $\mathcal{I}$ of cardinality $|\mathcal{I}|$. As in §1, the random variable T is exponentially-distributed with mean value λ, and represents the length of the contour. To ensure finiteness of the expressions that follow, we further assume $\lambda < \infty$. Sites or states within $\mathcal{I}$ will be denoted i and j. (Think of $\mathcal{I}$ as a discrete approximation to the state space $\mathcal{R} = \mathbb{R}^2 \times \mathbb{S}$ where a random process modeling each contour (in position and direction) takes values.) Let $\mathbb{1}\{condition\}$ denote the (indicator) function that takes on value 1 if *condition* is true, and the value 0 otherwise. With these notations we can define the *curve indicator random field* V *for a single curve* as:

$$V_i := \int_0^T \mathbb{1}\{R_t = i\}dt, \qquad \forall i \in \mathcal{I}.$$

Observe that V_i is the (random) amount of time that the Markov process spent in state i. In particular, V_i is zero unless the Markov process passed through site i. In the context of Brownian motion or other symmetric processes, V is variously known as the occupation measure or the local time of R_t [9, 10].

Generalizing to multiple curves, we pick a random number $\mathcal{N}$ and then choose $\mathcal{N}$ independent copies $R^{(1)}_{t_1}, \ldots, R^{(\mathcal{N})}_{t_\mathcal{N}}$ of the Markov process R_t, with independent lengths $T_1, \ldots, T_\mathcal{N}$, each distributed as T. To define the multiple curve CIRF, we take the superposition of the single-curve CIRFs $V^{(1)}, \ldots, V^{(\mathcal{N})}$ for the $\mathcal{N}$ curves.

Definition 1. *The curve indicator random field U is defined as:*

$$U_i := \sum_{n=1}^{\mathcal{N}} V_i^{(n)} = \sum_{n=1}^{\mathcal{N}} \int_0^{T_n} \mathbb{1}\{R_{t_n}^{(n)} = i\} dt_n, \qquad \forall i \in \mathcal{I}.$$

Thus U_i is the total amount of time that all of the Markov processes spent in site i. Again, observe that this definition satisfies our desiderata for an ideal edge/line map: (1) non-zero value where the contours are, and (2) zero-value elsewhere. Observe that the curve indicator random field really is a different object than the curves used to produce it. First, the CIRF describes a random *set* of curves, each one of which is a Markov process. Second, and more importantly, the CIRF is a stochastic function of *space*, a random field, whereas each curve is a random function of time. This transformation from a set of random curves to a random field makes the CIRF an idealization of local edge/line responses, and sets the stage for contour enhancement where the probability distribution of U will become our prior for inference. See Fig. 1 for some samples generated by the CIRF.

2.2 Statistics of the Curve Indicator Random Field

Probabilistic models in vision and pattern recognition have been specified in a number of ways. For example, Markov random field models [13] are specified via clique potentials and Gaussian models are specified via means and covariances. Here, instead of providing the distribution of the curve indicator random field itself, we report its moment generating functional, from which all moments are then computed.

Before doing so, we need to develop more Markov process theory. We first define the inner product $(a, b) := \sum_{i \in \mathcal{I}} a_i b_i$. The generator of the Markov process R_t is the $|\mathcal{I}| \times |\mathcal{I}|$ matrix $L = (l_{ij})$, and is the instantaneous rate of change of the probability transition matrix $P(t) = (p_{i,j})(t)$ for R_t. We assume that the Markov process is stationary, and so L does not vary with time t. Following Mumford [20], Williams and co-workers [24, 22], we use a discretization of the Markov process in position (x, y) and direction θ described in §1, having generator L:

$$L \approx \frac{\sigma^2}{2} \frac{\partial^2}{\partial \theta^2} - \cos\theta \frac{\partial}{\partial x} - \sin\theta \frac{\partial}{\partial y}.$$

To include the exponential distribution over lifetime (length) T of each particle, we construct a *killed* Markov process with generator $Q = L - \lambda^{-1} I$. (Formally, we do this by adding a single "death" state $\Cap$ to the discrete state space $\mathcal{I}$. When $t \geq T$, the process enters $\Cap$ and it cannot leave.) Slightly changing our notation, we shall now use R_t to mean the killed Markov process with generator Q. The Green's function matrix $G = (g_{ij})$ of the Markov process is the matrix $\int_0^\infty e^{Qt} dt = \int_0^\infty P(t) e^{-t/\lambda} dt$, where $P(t) = e^{Lt}$ (e^A denotes the matrix exponential of matrix A). The (i,j)-entry g_{ij} in the Green's function matrix represents the expected amount of time that the Markov process R_t spent in j (before death) given that the process started in i. The following is a well-known connection between the generator Q (a differential operator) and the Green's operator G (an integral operator):

Lemma 1. $G = -Q^{-1}$.

PROOF. We integrate from $t = 0$ to $t = \infty$ the standard result that $\frac{d}{dt} \mathrm{e}^{Qt} = Q e^{Qt}$: on the left side, we get $\mathrm{e}^{Qt} \,|_{t=0}^{\infty} = -I$; on the right, we get $Q \int_0^\infty \mathrm{e}^{Qt}\, dt = QG$. □

Although we are interested in the statistics of the general curve indicator random field U, we first consider the simpler, single-curve case, which we studied earlier in discrete-time [3]. The first step (Prop. 1) is to derive all the moments of the single-curve CIRF V. Then we shall summarize this result as a moment generating functional (Prop. 2). We use a *creation* vector $c \in \mathbb{R}^{|\mathcal{I}|}$ to act as a "probe" or "test function" function on V by taking an inner product:

$$(c, V) \;=\; \sum_i c_i \int \mathbb{1}\{R_t = i\} dt = \int \left(\sum_i c_i \, \mathbb{1}\{R_t = i\} \right) dt = \int c(R_t) dt, \qquad (2)$$

where $c(i) = c_i$. In the following, we let $\alpha := \lambda^{-1}$ to simplify expressions. We also introduce a final weighting $\nu(R_{T-})$ on the state of the curve just before death; ν can be used to encourage the curve to end in certain states over others.[1] Let $\mathbb{E}_i Z$ denote the expected value of the random variable Z given that $R_0 = i$, and let $\mathbb{E}_\mu Z$ denote the same expectation except given that $\mathbb{P}\{R_0 = i\} = \mu_i$. To reduce the clutter of many brackets we adopt the convention that the expectation operator applies to all multiplied (functions of) random variables to its

[1] Following [11], we use the notation $T-$ to represent the left limit approaching T from below, i.e., $\nu(R_{T-}) = \lim_{t \nearrow T} \nu(R_t)$.

right: e.g., $\mathbb{E}f(X)g(Y) := \mathbb{E}[f(X)g(Y)]$. We now prove a formula for the moments of (c, V), a "probed" single-curve CIRF V.

Proposition 1. *The k-th moment of (c, V) with initial distribution μ and final weighting $\nu = \nu(i) = \nu_i, i \in \mathcal{I}$ is:*

$$\mathbb{E}_\mu(c, V)^k \nu(R_{T-}) = \alpha k!(\mu, (GC)^k G\nu). \tag{3}$$

PROOF. We first consider the case where $\mu_j = \delta_{i,j}$, and then generalize. Recall the formula for exponentially-distributed length T: $\mathbb{P}\{T = t\} = \alpha \, \mathrm{e}^{-\alpha t}$. Substituting this and (2) into the left side of (3), we get:

$$\begin{aligned} \mathbb{E}_i(c, V)^k \nu(R_{T-}) &= \mathbb{E}_i \left(\int_0^T c(R_t) dt \right)^k \nu(R_{T-}) \\ &= \alpha \mathbb{E}_i \int_0^\infty \mathrm{e}^{-\alpha t} \left(\int_0^t c(R_{t'}) dt' \right)^k \nu(R_t) dt, \end{aligned} \tag{4}$$

where we have used the fact that:

$$\int f(t-) dt = \int f(t) dt,$$

for piecewise continuous functions f. We further note that:

$$\begin{aligned} \left(\int_0^t c(R_{t'}) dt' \right)^k &= \int_0^t \cdots \int_0^t c(R_{t_1}) \cdots c(R_{t_k}) dt_1 \cdots dt_k \\ &= k! \int \cdots \int_{0 \le t_1 \le \cdots \le t_k \le t} c(R_{t_1}) \cdots c(R_{t_k}) dt_1 \cdots dt_k, \end{aligned} \tag{5}$$

where the second line follows because there are $k!$ orthants in the k-dimensional cube $[0, t]^k$, each having the same integral by relabeling the t_i's appropriately. Taking integrals iteratively starting with respect to t_1, the right side of (5) becomes by induction in k:

$$k! \int_0^t \int_0^{t_k} \cdots \int_0^{t_2} c(R_{t_1}) \cdots c(R_{t_k}) dt_1 \cdots dt_k.$$

The right side of (4) then becomes:

$$\alpha k! \int_0^\infty \mathrm{e}^{-\alpha t} \int_0^t \int_0^{t_k} \cdots \int_0^{t_2} \sum_{i_1, \ldots, i_k, j} \mathbb{P}_i\{R_{t_1} = i_1, \ldots, R_{t_k} = i_k, R_t = j\} \cdot c(i_1) \cdots c(i_k) \nu(j) dt_1 \cdots dt_k$$

$$= \alpha k! \sum_{i_1,\ldots,i_k,j} c(i_1)\cdots c(i_k)\nu(j)\Big[\int_0^\infty \mathrm{e}^{-\alpha t} \cdot \Big\{\int_0^t \int_0^{t_k} \cdots \int_0^{t_2} p_{i,i_1}(t_1)p_{i_1,i_2}(t_2-t_1) \cdots p_{i_{k-1},i_k}(t_k-t_{k-1})p_{i_k,j}(t-t_k)dt_1dt_2\cdots dt_{k-1}\Big\}dt_k\Big], \quad (6)$$

using the Markovianity and stationarity of R_t. Recalling that the formula for the convolution of two functions f and g is:

$$(f*g)(t) = \int_0^t f(\tau)g(t-\tau)d\tau,$$

we see that the expression in braces in (6) can be written as:

$$\begin{aligned}
&\int_0^t \cdots \int_0^{t_2} p_{i,i_1}(t_1)p_{i_1,i_2}(t_2-t_1)dt_1\cdots p_{i_k,j}(t-t_k)dt_{k-1} \\
&= \int_0^t \cdots \int_0^{t_3} (p_{i,i_1} * p_{i_1,i_2})(t_2)p_{i_2,i_3}(t_3-t_2)dt_2\cdots p_{i_k,j}(t-t_k)dt_{k-1} \\
&= (p_{i,i_1} * p_{i_1,i_2} * \cdots p_{i_{k-1},i_k} * p_{i_k,j})(t),
\end{aligned}$$

which is a k-fold convolution by induction in k. Now observe that the expression in brackets in (6) is the Laplace transform $\mathcal{L}\{h(t)\}(\alpha)$ with respect to t, evaluated at α, of the expression in braces, say $h(t)$. Therefore by using the convolution rule of the Laplace transform k times, the expression in brackets in (6) becomes:

$$\mathcal{L}\{p_{i,i_1}\}(\alpha)\cdot\mathcal{L}\{p_{i_1,i_2}\}(\alpha)\cdot\ \cdots\ \cdot\mathcal{L}\{p_{i_{k-1},i_k}\}(\alpha)\cdot\mathcal{L}\{p_{i_k,j}\}(\alpha).$$

But as shown earlier we know that $g_{i,j} = \int_0^\infty \mathrm{e}^{-\alpha t}\, p_{i,j}(t)dt = \mathcal{L}\{p_{i,j}\}(\alpha)$, and so we can write:

$$\begin{aligned}
&\mathbb{E}_i(c,V)^k\nu(R_{T-}) \\
&= \alpha k! \sum_{i_1,\ldots,i_k,j} c(i_1)\cdots c(i_k)\nu(j)g_{i,i_1}g_{i_1,i_2}\cdots g_{i_{k-1},i_k}g_{i_k,j} \qquad (7)\\
&= \alpha k! \sum_{i_1} g_{i,i_1}c(i_1)\left\{\sum_{i_2} g_{i_1,i_2}c(i_2)\cdots\left[\sum_{i_k} g_{i_{k-1},i_k}c(i_k)\left(\sum_j g_{i_k,j}\nu(j)\right)\right]\cdots\right\} \\
&= \alpha k! \sum_{i_1} (GC)_{i,i_1}\left\{\sum_{i_2}(GC)_{i_1,i_2}\cdots\left[\sum_{i_k}(GC)_{i_{k-1},i_k}(G\nu)_{i_k}\right]\cdots\right\} \\
&= \alpha k!((GC)^kG\nu)_i.
\end{aligned}$$

Since for any random variable Z we have $\mathbb{E}_\mu Z = \sum_i \mathbb{P}\{R_0=i\}\mathbb{E}[Z|R_0=i] = \sum_i \mu_i\mathbb{E}_i Z$, the result follows. □

We now compute the moment generating functional for the single-curve case using Prop. 1. This is known as the Feynman-Kac formula [9]. First we define the Green's function matrix $G(c)$ with spatially-varying "creation" c as the Green's function matrix for the killed Markov process with extra killing $-c$, i.e., having generator $Q(c) := Q + C$, where $C = \mathrm{diag}(c_1, \ldots, c_{|\mathcal{I}|})$. We use the term creation for c because it behaves exactly opposite to the decay or death term λ^{-1} in $Q = L - \lambda^{-1} I$. Using an argument similar to the proof of Lemma 1, $G(c) := -Q(c)^{-1} = -(Q + C)^{-1}$. We provide an explicit condition for the invertibility of $Q + C$ later, in Prop. 3.

Proposition 2. *For all $c \in \mathbb{R}^{|\mathcal{I}|}$ such that $|c|$ is sufficiently small,*

$$\mathbb{E}_\mu \exp(c, V)\nu(R_{T-}) = \alpha(\mu, G(c)\nu).$$

PROOF. Using the power series representation of the exponential, we write:

$$\begin{aligned} \mathbb{E}_\mu \exp(c, V)\nu(R_{T-}) &= \mathbb{E}_\mu \sum_{k=0}^{\infty} \frac{(c, V)^k}{k!} \nu(R_{T-}) \\ &= \sum_{k=0}^{\infty} \mathbb{E}_\mu (c, V)^k \nu(R_{T-})/k! = \sum_{k=0}^{\infty} \alpha(\mu, (GC)^k G\nu), \end{aligned} \tag{8}$$

using Prop. 1. Recalling the fact that $\sum_{k=0}^{\infty} A^k = (I - A)^{-1}$, as long as $||A|| < 1$, the right hand side of (8) becomes:

$$\alpha(\mu, \left[\sum_{k=0}^{\infty} (GC)^k\right] G\nu) = \alpha(\mu, (I - GC)^{-1} G\nu),$$

as long as (for some operator norm) $||GC|| < 1$. Since $C = \mathrm{diag}(c_1, \ldots, c_{|\mathcal{I}|})$ is diagonal, the matrix GC is simply G with the i-th column weighted by c_i. Therefore $||GC|| < 1$ for $|c|$ sufficiently small. The result follows because $(I - GC)^{-1} = (Q^{-1}Q + Q^{-1}C)^{-1} = -(Q + C)^{-1} G^{-1}$. □

Observe that to evaluate the Feynman-Kac formula, one must solve the linear system $(Q + C)h + \nu = 0$ for h. This equation will become a key component in the filter of §4. The proof of Prop. 2 suggests with the following condition for the convergence of the Feynman-Kac formula first studied by Khas'minskii [14].

Proposition 3. *The moment generating functional of the (single-curve) curve indicator random field V converges if $||G|c|\,||_\infty < 1$.*

PROOF. Using the ∞-norm, we consider the convergence condition in the proof of Prop. 2:

$$1 > ||GC||_\infty = \sup_i \sum_j |(GC)_{i,j}| = \sup_i (G|C|\mathbf{1})_i = \sup_i (G|c|)_i,$$

where $\mathbf{1} = (1, \ldots, 1)$ and using the facts that G has positive entries and C is diagonal. □

The Khas'minskii condition is easy to check: one just takes the componentwise absolute value of c (which later will be an input), and then "blur" it with the Green's operator G. If any component of the result is greater than 1, the moment generating functional may not converge. As we shall see in §4, this can be interpreted as a kind of stability condition as well as constraining our generative model.

As the interpretation of the "final weighting" ν above may seem mysterious, we now restrict μ and ν to be finite measures satisfying the normalization constraint $1 = (\mu, G\nu)$. (If this equality is not satisfied, one need only divide by a suitable normalizing constant.)

Corollary 1. *Suppose that the joint distribution over initial and final positions is $\mathbb{P}\{R_0 = i, R_{T-} = j\} = \mu_i g_{i,j} \nu_j$. Then the moment generating functional of V, with this distribution over initial and final states, is:*

$$\mathbb{E} \exp(c, V) = (\mu, G(c)\nu). \tag{9}$$

Although not studied here, it is interesting to consider the problem of finding those measures μ, ν that induce a $\mathbb{P}\{R_0, R_{T-}\}$ with prescribed marginals over initial and final states. Before we begin the proof of this corollary, we state a basic result that will also be used later.

Lemma 2. *If X and Z are random variables, and X is discrete (i.e., X can only take on one of an at most countable number of values x), then:*

$$\mathbb{E}[Z|X = x] = \frac{\mathbb{E} Z \, \mathbb{1}\{X = x\}}{\mathbb{P}\{X = x\}}.$$

PROOF. We compute:

$$\mathbb{E}[Z|X=x] = \int z\mathbb{P}\{Z \in dz|X=x\} = \int z\mathbb{P}\{Z \in dz, X=x\}/\mathbb{P}\{X=x\}$$
$$= \int \sum_{x'} z\,\mathbb{1}\{x'=x\}\mathbb{P}\{Z \in dz, X=x'\}/\mathbb{P}\{X=x\} = \mathbb{E}Z\,\mathbb{1}\{X=x\}/\mathbb{P}\{X=x\}.$$

□

PROOF OF COROLLARY 1. Observe that $\mu_i g_{i,j}\nu_j$ is indeed a distribution. We now compute the result:

$$\begin{aligned}
\mathbb{E}\exp(c,V) &= \sum_{i,j} \mathbb{P}\{R_0=i, R_{T_-}=j\}\mathbb{E}[\exp(c,V)|R_0=i, R_{T_-}=j] \\
&= \sum_{i,j} \mu_i g_{i,j}\nu_j \mathbb{E}_i[\exp(c,V)|R_{T_-}=j] \\
&= \sum_{i,j} \mu_i g_{i,j}\nu_j \mathbb{E}_i[\exp(c,V)\,\mathbb{1}\{R_{T_-}=j\}]/\mathbb{P}_i\{R_{T_-}=j\},
\end{aligned}$$

using Lemma 2. But using Prop. 2 we see that:

$$\mathbb{P}_i\{R_{T_-}=j\} = \mathbb{E}_i \exp(0,V)\,\mathbb{1}\{R_{T_-}=j\} = \alpha g_{i,j},$$

and therefore:

$$\begin{aligned}
\mathbb{E}\exp(c,V) &= \alpha^{-1}\sum_{i,j} \mu_i\nu_j \mathbb{E}_i[\exp(c,V)\,\mathbb{1}\{R_{T_-}=j\}] \\
&= \alpha^{-1}\mathbb{E}_\mu\left[\exp(c,V)\left(\sum_j \nu_j\,\mathbb{1}\{R_{T_-}=j\}\right)\right] \\
&= \alpha^{-1}\mathbb{E}_\mu\left[\exp(c,V)\nu(R_{T_-})\right].
\end{aligned}$$

The result follows after another application of Prop. 2 to the above expectation. □

The next corollary shows all of the (joint) moments of V. In our earlier work in discrete time [3], instead of computing an expectation over $(c,V)^k$, we in effect computed an expectation over $(c^{(1)},V)\cdots(c^{(k)},V)$, for arbitrary vectors $c^{(1)},\ldots,c^{(k)}$. In observing the connection to the Feynman-Kac formula, the proofs were simplified and summarized succinctly in the moment generating functional. In addition, the weighting over final states (the other end of the contour) was included. Let perm_k denote the set of permutations of the integers $1,\ldots,k$.

Corollary 2. *If $k \geq 1$, the k-th (joint) moment of V at sites $i_1, \ldots, i_k$ is:*

$$\mathbb{E}\, V_{i_1} \cdots V_{i_k} = \sum_{i,j} \mu_i \nu_j \sum_{a \in \mathrm{perm}_k} g_{i i_{a_1}} g_{i_{a_1} i_{a_2}} \cdots g_{i_{a_{k-1}} i_{a_k}} g_{i_{a_k} j}. \tag{10}$$

PROOF. Take partial derivatives of (9) with respect to $c_{i_1}, \ldots, c_{i_k}$. The only nonzero terms come from differentiating an expression proportional to (7). □

3 Multiple curve moment generating functional

In order to model more than one curve in an image, we need a joint distribution over both the number of curves and the curves (and the corresponding CIRFs) themselves. To make our computations concrete, we adopt a Poisson distribution over the number $\mathcal{N}$ of curves, and assume conditional independence of the curves given $\mathcal{N}$. To compute the moment generating functional of this (multiple-curve) CIRF as a Poisson distribution over (single-curve) CIRFs, we first consider the general case of Poisson "point" process given a distribution over each point, where point will be interpreted as an entire single-curve CIRF.

3.1 The Poisson measure construction

We begin with a finite measure[2] $P : \mathcal{F} \to \mathbb{R}_+$ over the measure space $(\Omega, \mathcal{F})$, where $\mathcal{F}$ is a σ-algebra. Intuitively, the finite measure P is the (unnormalized) distribution over "points" $\omega \in \Omega$, where in this paper ω is a single-curve CIRF realization (i.e., a curve image) and Ω is the set of single-curve CIRF realizations (i.e., all possible curve images). We shall now define a probability distribution over random configurations $\boldsymbol{\omega} = (\omega_1, ..., \omega_{\mathcal{N}}) \in \mathrm{Con}(\Omega) := \{\Omega^0 = \varnothing, \Omega^1 = \Omega, \Omega^2 = \Omega \times \Omega, \Omega^3, \ldots\}$, where each ω_n is a curve in Ω and $\mathcal{N}$ is the random number of curves. In our context, Ω^0 is the 0-curve configuration, Ω^1 is the one-curve configuration, and so on. We now compute the *Poisson point measure* via its expectation $\mathbb{E}F$ on any (measurable) function $F : \mathrm{Con}(\Omega) \to \mathbb{R}$ (clearly this defines a probability distribution for we could take F as an indicator over any (measurable) subset of $\mathrm{Con}(\Omega)$ to get its probability).

[2] A finite measure can always be normalized to a probability distribution because $P(\Omega) < \infty$. In particular, $\tilde{P}(\omega) := P(\omega)/P(\Omega)$ is a (normalized) probability distribution over ω.

Proposition 4. *Suppose $\mathcal{N}$ is a Poisson deviate with mean $P(\Omega)$. Further suppose that the points $\omega_1, \ldots, \omega_n$ are (conditionally) independent and identically distributed with $P(\cdot)/P(\Omega)$, given $\mathcal{N} = n$. Then:*

$$\mathbb{E}F := \sum_{n=0}^{\infty} \frac{e^{-P(\Omega)}}{n!} \int_{\Omega^n} F(\omega_1, \ldots, \omega_n) P(d\omega_1) \cdots P(d\omega_n). \tag{11}$$

PROOF. We need only take a conditional expectation and recall the formula for the Poisson distribution, as follows:

$$\begin{aligned}
\mathbb{E}F &= \mathbb{E}(\mathbb{E}[F|\mathcal{N}]) \\
&= \mathbb{E}(\mathbb{E}F(\omega_1, \ldots, \omega_{\mathcal{N}})) \\
&= \mathbb{E}\left(\int_{\Omega^{\mathcal{N}}} F(\omega_1, \ldots, \omega_{\mathcal{N}}) \left(P(d\omega_1)/P(\Omega)\right) \cdots \left(P(d\omega_{\mathcal{N}})/P(\Omega)\right) \right) \\
&= \sum_{n=0}^{\infty} \frac{\mathrm{e}^{-P(\Omega)}\, P(\Omega)^n}{n!} \left(P(\Omega)^{-n} \int_{\Omega^{\mathcal{N}}} F(\omega_1, \ldots, \omega_{\mathcal{N}}) P(d\omega_1) \cdots P(d\omega_{\mathcal{N}}) \right).
\end{aligned}$$

The result follows. □

The above presentation of the Poisson point measure is based on Dynkin [9].

3.2 Application to the Curve Indicator Random Field

We now consider the joint distribution over many curves that we sought. Suppose there are $\bar{\mathcal{N}}$ contours on average, and that μ and ν are finite measures on $\mathcal{I}$ (vectors in $\mathbb{R}^{|\mathcal{I}|}$), characterizing the initial and final positions, respectively, of the Markov processes $\{R_t^{(n)} : n = 1, \ldots, \mathcal{N}\}$. As before, these measures satisfy the normalization constraint $(\mu, G\nu) = 1$. For general-purpose contour enhancement, we typically have no a-priori preference for the start and end locations of each contour, and so we set these measures proportional to the constant vector $\mathbf{1} = (1, \ldots, 1)$. One can show that by letting $\mu_i = |\mathcal{I}|^{-1}, \nu_i = \lambda^{-1}, \forall i \in \mathcal{I}$, the above constraint is satisfied. We now state and prove the following key theoretical result of this paper, which is most closely related to the work by Dynkin [9].

Proposition 5. *The moment generating functional of the curve indicator random field U is:*

$$\mathbb{E}\exp(c, U) = \exp(\mu, \bar{\mathcal{N}}(G(c) - G)\nu).$$

PROOF. To take advantage of the Poisson point measure construction, we let ω be a single-curve CIRF for the killed Markov process $R_t, t \in [0, T-)$, such that the finite measure $P = P(\omega)$ is the probability distribution for ω but multiplied by the constant $\bar{\mathcal{N}}$, i.e., $P(\Omega) = \bar{\mathcal{N}}$. Let $F := \exp(c, U) = \exp \sum_{n=0}^{\mathcal{N}} (c, V^{(n)}) = \prod_{n=0}^{\mathcal{N}} \exp(c, V^{(n)})$, where $V^{(n)}$ is a function of ω_n. Applying (11) we obtain:

$$\begin{aligned}
\mathbb{E} \exp(c, U) &= \sum_{n'=0}^{\infty} \frac{\mathrm{e}^{-P(\Omega)}}{n'!} \int_{\Omega^{n'}} \prod_{n=0}^{n'} \exp(c, V^{(n)}) P(d\omega_1) \cdots P(d\omega_{n'}) \\
&= \sum_{n'=0}^{\infty} \frac{\mathrm{e}^{-P(\Omega)}}{n'!} \prod_{n=0}^{n'} \left(\int_{\Omega} \exp(c, V^{(n)}) P(d\omega_n) \right) \\
&= \mathrm{e}^{-P(\Omega)} \sum_{n'=0}^{\infty} \frac{1}{n'!} \prod_{n=0}^{n'} \left(\int_{\Omega} \exp(c, V^{(1)}) P(d\omega_1) \right),
\end{aligned}$$

since $V^{(1)}, \ldots, V^{(n')}$ are identically distributed. But then the latter integral is $\bar{\mathcal{N}} \mathbb{E} \exp(c, V)$, and so the above sum becomes:

$$\sum_{n'=0}^{\infty} \left(\bar{\mathcal{N}} \mathbb{E} \exp(c, V) \right)^{n'} / n'! = \exp(\bar{\mathcal{N}} \mathbb{E} \exp(c, V)).$$

So using $P(\Omega) = \bar{\mathcal{N}}(\mu, G(\mathbf{0})\nu)$ and Prop. 1, we conclude:

$$\mathbb{E} \exp(c, U) = \exp(\bar{\mathcal{N}}(\mu, G(c)\nu) - \bar{\mathcal{N}}(\mu, G\nu)).$$

□

While this result may seem abstract, it is actually very useful. First observe that how similar in form it is to the single-curve case. More importantly, with Prop. 5 we can compute the higher-order cumulants [21] of U (recall that the moments define the cumulants and vice versa):

Corollary 3. *If $k \geq 1$, the k-th (joint) cumulant of the curve indicator random field U at sites $i_1, \ldots, i_k$ is:*

$$\mathrm{cum}\{U_{i_1}, \ldots, U_{i_k}\} = \bar{\mathcal{N}} \sum_{i,j} \mu_i \nu_j \sum_{a \in \mathrm{perm}_k} g_{i i_{a_1}} g_{i_{a_1} i_{a_2}} \cdots g_{i_{a_{k-1}} i_{a_k}} g_{i_{a_k} j}. \tag{12}$$

PROOF. Since the cumulant generating functional of U, which is the natural logarithm of the moment generating functional of U, differs from the moment generating functional of V by an additive constant, we use (10) with no further work. □

The cumulant formula has a simple interpretation. First recall that the Green's function g_{ij} is the expected amount of time spent by R_t in state j given that it started in i. For any ordering of the k points we take the product of the g_{ij}'s for the successive points in order (the first and last factors deal with the initial and final points). Since the contour could have passed through the points in any order, all must be considered.

We can rephrase Cor. 3 to show the mean and covariance of the CIRF [2]. Let G^* denote the transpose of G.

Corollary 4. *Suppose that* $\mu_i = |\mathcal{I}|^{-1}, \nu_i = \lambda^{-1}, \forall i \in \mathcal{I}$. *The mean of the curve indicator random field* U *is* $\mathbb{E}\, U_i = \bar{\mathcal{N}}\lambda|\mathcal{I}|^{-1}, \forall i \in \mathcal{I}$. *The covariance matrix of* U *is* $\operatorname{cov} U = \bar{\mathcal{N}}\lambda|\mathcal{I}|^{-1}(G + G^*)$.

Several "columns" of the covariance matrix for the curvature process are illustrated in Fig. 2, by taking its impulse response for several positions, directions and curvatures.

Note that by the cumulants of U of order greater than two are generally *not* zero, which shows that *the curve indicator random field is non-Gaussian.* Despite that, its moment generating functional has a tractable form that we shall directly exploit next.

4 Minimum Mean Square Error Filtering

Instead of the unknown random field U, what we actually observe is a random field M of (edge or line) measurements. Given a realization m of M, we seek to find that approximation $\tilde{u}$ of U that minimizes the mean square error (MMSE):

$$\tilde{u} := \arg\min_u \mathbb{E}_m ||u - U||^2,$$

where $\mathbb{E}_m$ the denotes taking an expectation conditioned on the measurement realization m. It is well-known that the posterior mean is the MMSE estimate:

$$\tilde{u} = \mathbb{E}_m U,$$

but in many interesting, non-Gaussian, cases this is extremely difficult to compute. In our context, however, we are fortunate to be able to make use of the moment generating functional to simplify computations.

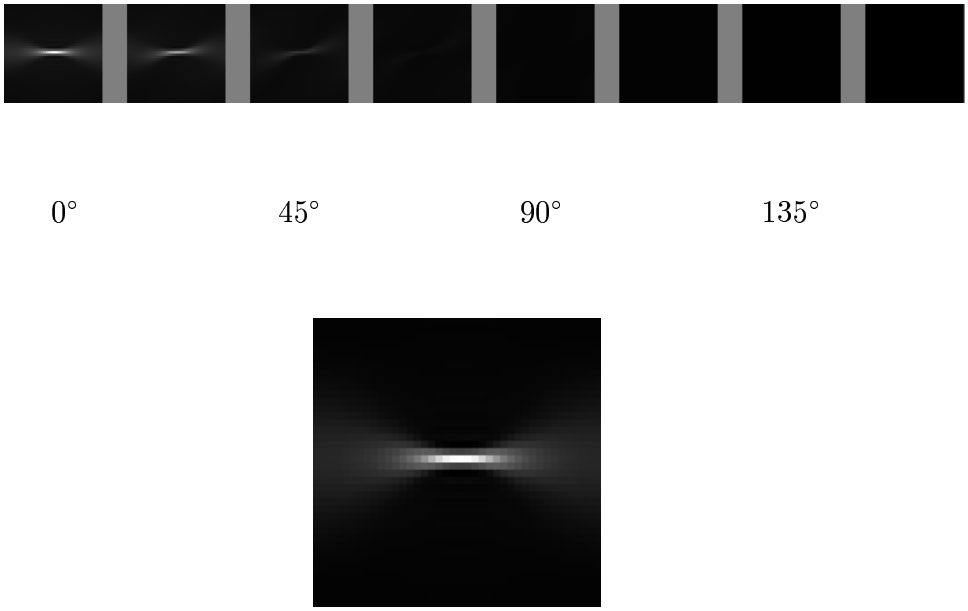

Figure 2: Impulse response of the covariance matrix (Cor. 4) for the curve indicator random field for Mumford's Markov process in (x, y, θ). Impulse is located at image center, with direction 0°, then Gaussian blurred with radius 1.0. Parameters are $\sigma = 1/8, \lambda = 200, \bar{\mathcal{N}} = 10$. (Top) Stack of images, arranged left to right, each for a different direction. (Bottom) Sum over θ of stack above, producing single image in (x, y). Observe the elongation of the response along the horizontal direction, capturing the smoothness and length of the contours.

Before developing our MMSE estimator, we must define our likelihood function $p(M|U)$. First let H_i be the binary random variable taking the value 1 if one of the contours passed through (or "hit") site i, and 0 otherwise, and so H is a binary random field on $\mathcal{I}$. In this paper we consider conditionally independent, local likelihoods: $p(M|H) = p(M_1|H_1) \cdots p(M_{|\mathcal{I}|}|H_{|\mathcal{I}|})$. Following [12, 25], we consider two distributions over measurements at site i: $p_{\text{on}}(M_i) := p(M_i|H_i = 1)$ and $p_{\text{off}}(M_i) := p(M_i|H_i = 0)$. It follows [25] that $\ln p(M|H) = \sum_i H_i \ln \frac{p_{\text{on}}(M_i)}{p_{\text{off}}(M_i)}$. Now let τ be the average amount of time that the Markov process spends in a site, given that the process hit the site, and observe that U_i/τ and H_i are equal on average. This suggests that we replace H with U/τ above to generate a likelihood in U:

$$\ln p(M|U) \approx \sum_i c_i U_i = (c, U), \text{ where } c_i = c_i(M_i) = \tau^{-1} \ln \frac{p_{\text{on}}(M_i)}{p_{\text{off}}(M_i)}.$$

As shown in [1], for uniform initial and final weightings μ and ν the posterior mean in this

case becomes:

$$\mathbb{E}_m U_i \approx \frac{\partial}{\partial c_i}(\mathbb{E}\exp(c,U))(c) = f_i b_i, \tag{13}$$

where $f = (f_1, \ldots, f_{|\mathcal{I}|})$ is the solution to the forward equation:

$$(Q + C)f + \gamma = 0$$

and $b = (b_1, \ldots, b_{|\mathcal{I}|})$ is the solution to the backward equation:

$$(Q^* + C)b + \gamma = 0,$$

where γ is a constant. Note that these equation can be viewed as partial differential equations, if we return from the discrete state space $\mathcal{I}$ to a continuum, such as $\mathbb{R}^2 \times \mathbb{S}$.

Observe that two nonlinearities arise in this posterior mean. First, we are taking the product of the forward and backward solutions.[3] Second, although both the forward and backward equations are linear, notice that the input c does not enter via the usual forcing term on the right hand side. Rather, the input arises as a modification along the diagonal of the linear operator; for example, $f = (I - GC)^{-1}G\gamma = \sum_{k=0}^{\infty}(GC)^k G\gamma$, since the Green's operator $G = -Q^{-1}$. Because this sum has terms polynomial in C for all orders k, we have a (nonlinear) Volterra filter.

To illustrate how this filter works, we again return to the (x, y, θ)-space of positions and directions where Mumford's Markov process (1) takes values. An ideal contour might be the vertical line (with local direction 90°) shown at the top Fig. 3, depicted as a stack of images with increasing direction, from left to right. Gaussian noise was added added (SNR=7.96) and, for numerical reasons outside the scope of this paper, the result was Gaussian blurred with a radius of 1., producing the image stack (the measurement field M) in the middle. The goal was to restore the noisy stack to the original using (13), setting c proportional to M. The constant of proportionality was the largest satisfying the convergence constraint of Prop. 3; larger constants can give erratic results and so Prop. 3 provides a very practical guideline. Parameters were $\sigma = .1, \lambda = 100, \bar{\mathcal{N}} = 1$. The result (bottom) shows the that the noise has mostly been removed and the contour was enhanced. Also observe the faint streakiness having the same orientation of each θ-slice in the stack: filtering is mostly in

[3] This is analogous to the source/sink product in the stochastic completion field [24].

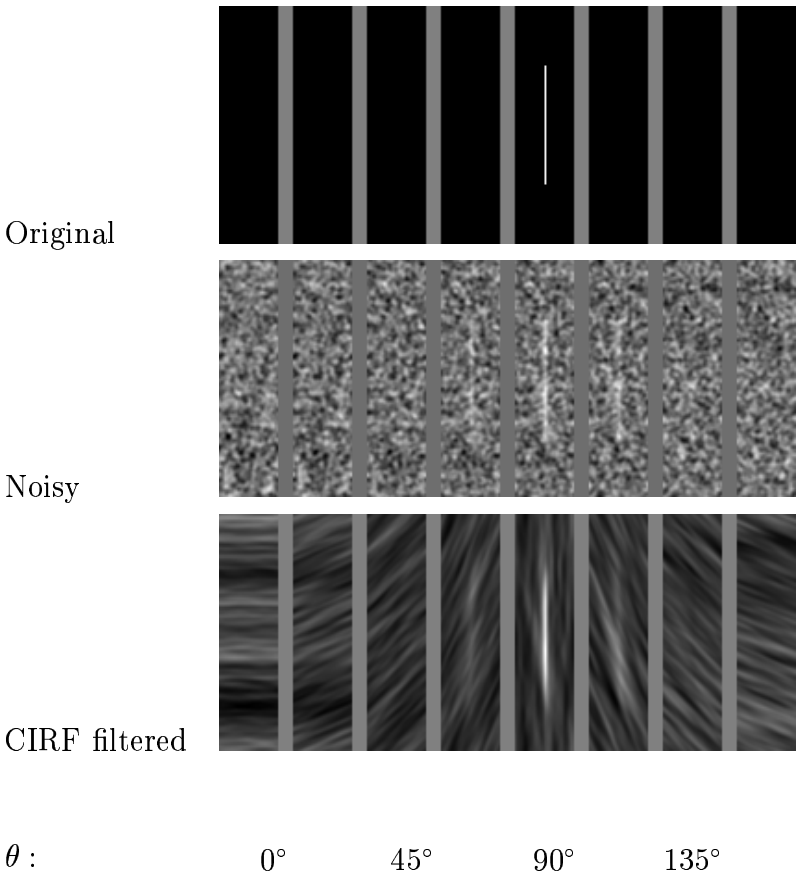

Figure 3: Contour filtering via the CIRF via (13).

straight lines, as suggested by the covariance impulse response in Fig. 2. More elaborate computations on medical images will be presented in [1].

5 Conclusion

The purpose of this paper was to present an exact, generative model for image contours, called the curve indicator random field, and to derive its joint statistics. The analysis was successful in providing an explicit formula for all cumulants and a simple expression for the moment generating functional, even though the model is non-Gaussian. We also presented a filter that exploits the model for enhancing oriented edge maps. Curiously, the moment generating functional result requires that a convergence condition be satisfied, a condition which can be explicitly checked to ensure the "stability" of our filter.

Acknowledgements

We would like to thank Patrick Huggins for carefully reading the initial manuscript and he and Athinodoros Georghiades for the proof of Lemma 1. We also thank Steven Haker for stimulating discussions. Financial support was provided by AFOSR.

References

[1] J. August. *The Curve Indicator Random Field*. PhD thesis, Yale University, 2001.

[2] J. August and S. W. Zucker. The curve indicator random field: curve organization via edge correlation. In K. Boyer and S. Sarkar, editors, *Perceptual Organization for Artificial Vision Systems*, pages 265–288. Kluwer Academic, Boston, 2000.

[3] J. August and S. W. Zucker. The moments of the curve indicator random field. In *Proceedings of the 2000 Conference on Information Sciences and Systems*, volume 1, pages WP5–19–WP5–24,, Princeton, NJ, March 2000.

[4] J. August and S. W. Zucker. A markov process using curvature for filtering curve images. In *Energy Minimization Methods for Computer Vision and Pattern Recognition*, 2001.

[5] J. Besag. On the statistical analysis of dirty pictures. *J. R. Statist. Soc. B*, 48(3):259–302, 1986.

[6] R. Chellapa. Two-dimensional discrete gaussian markov random field models for image processing. In *Progress in Pattern Recognition*, pages 79–112.

[7] I. J. Cox, J. M. Rehg, and S. Hingorani. A Bayesian Multiple-Hypothesis Approach to Edge Grouping and Contour Segmentation. *International Journal of Computer Vision*, 11:5–24, 1993.

[8] G. C. Cross and A. K. Jain. Markov random field texture models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 5(1):25–39, 1983.

[9] E. B. Dynkin. Markov processes as a tool in field theory. *Journal of Functional Analysis*, 50:167–187, 1983.

[10] E. B. Dynkin. Gaussian and non-gaussian fields associated with markov processes. *Journal of Functional Analysis*, 55:344–376, 1984.

[11] P. J. Fitzsimmons and J. Pitman. Kac's moment formula and the feynman-kac formula for additive functionals of a markov process. *Stochastic Processes and their Applications*, 79:117–134, 1999.

[12] D. Geman and B. Jedynak. An active testing model for tracking roads in satellite images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(1):1–14, 1996.

[13] S. Geman and D. Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6):721–741, 1984.

[14] R. Z. Khas'minskii. On positive solutions of the equation uu+vu. *Theory of Probability and Its Applications*, 4(3):309–318, 1959.

[15] R. L. Lagendijk and J. Biemond. *Iterative Identification and Restoration of Images*. Kluwer, Boston, 1991.

[16] V. A. F. Lamme. The neurophysiology of figure-ground segregation in primary visual cortex. *Journal of Neuroscience*, 15(2):1605–1615, 1995.

[17] J. L. Marroquin. A markovian random field of piecewise straight lines. *Biological Cybernetics*, 61:457–465, 1989.

[18] A. Martelli. An application of heuristic search methods to edge and contour detection. *Comm. of the ACM*, 19(2):73–83, 1976.

[19] U. Montanari. On the optimum dection of curves in noisy pictures. *Comm. ACM*, 14:335–345, 1971.

[20] D. Mumford. *Algebraic Geometry and Its Applications*, chapter Elastica and Computer Vision, pages 491–506. Springer-Verlag, 1994.

[21] C. L. Nikias and A. P. Petropulu. *Higher-Order Spectra Analysis: A Nonlinear Signal Processing Framework*. Prentice Hall, Englewood Cliffs, NJ, 1993.

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

[23] M. Turk and A. Pentland. Eigenfaces for recognition. *J. Cognitive Neuroscience*, 3(1):71–96, 1991.

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

[25] A. L. Yuille and J. M. Coughlan. Fundamental bounds of bayesian inference: Order parameters and phase transitions for road tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(2):160–173, 2000.

[26] S. C. Zhu and D. Mumford. Prior learning and gibbs reaction-diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(11), Nov. 1997.