Steerable Random Fields

Stefan Roth
GRIS
TU Darmstadt

Michael J. Black
Dept. of Computer Science
Brown University

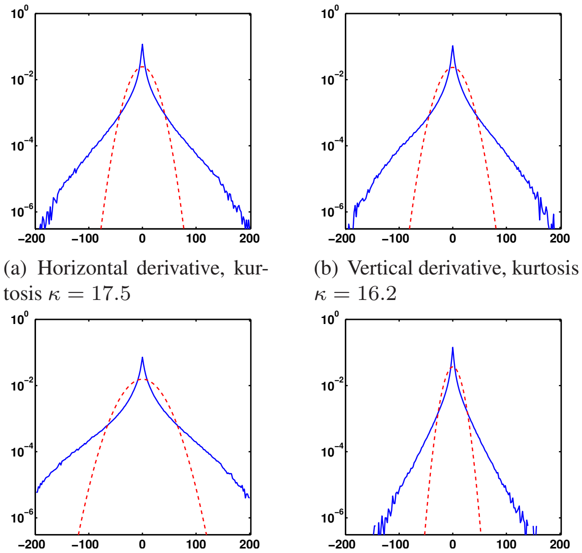
Abstract

In contrast to traditional Markov random field (MRF) models, we develop a Steerable Random Field (SRF) in which the field potentials are defined in terms of filter responses that are steered *to the local image structure. In particular, we use the structure tensor to obtain derivative responses that are either aligned with, or orthogonal to, the predominant local image structure, and analyze the statistics of these steered filter responses in natural images. Clique potentials are defined over steered filter responses using a Gaussian scale mixture model and are learned from training data. The SRF model connects random field models with anisotropic regularization and provides a statistical motivation for the latter. We demonstrate that steering the random field to the local image structure improves image denoising and inpainting performance compared with traditional pairwise MRFs.*

1. Introduction

Random field (RF) models have a long history in computer vision, particularly in problems of low-level vision such as image restoration [4, 13], stereo reconstruction, and optical flow estimation. In these applications, Markov random fields (MRFs) perform spatial (or spatio-temporal) regularization by imposing prior knowledge on the types of admissible images, depth maps, flow fields, *etc*. Discriminatively trained MRFs (*i.e*., conditional random fields (CRFs)) have also been used to directly model the posterior distribution in a number of vision problems, *e.g*., [6, 10]. In general, MRF models are receiving renewed attention due to improvements in learning and inference algorithms over the last few years.

In most cases, however, MRFs have been limited in three different regards: *(1)* They have used very simple neighborhood structures. Most models in low-level vision are based on pairwise graphs, where the potential functions are formulated in terms of pixel differences (image derivatives) between neighboring pixels; *(2)* In many cases, potentials have remained hand-defined and hand-tuned. Consequently, many MRFs do not necessarily reflect the statistical properties of the data; *(3)* MRF models have typically not been spatially adaptive, *i.e*., their potentials do not depend on the spatial location within the image. The first two shortcomings have been addressed by a number of recent approaches [12, 18, 29]. In this paper we go beyond previous work by addressing the third limitation as well, and develop a spatially adaptive random field model.



(a) Horizontal derivative, kurtosis $\kappa = 17.5$

(b) Vertical derivative, kurtosis $\kappa = 16.2$

(c) Orthogonal derivative, kurtosis $\kappa = 12.7$

(d) Aligned derivative, kurtosis $\kappa = 16.2$

Figure 1. **Marginal filter response statistics (log scale)** of standard derivatives (top row) and steered derivatives (bottom row). The histograms are shown in solid blue; a fit with a Gaussian is shown in dashed red. Note the difference between the orthogonal and the aligned case of the steered derivatives.

In particular, we develop a Steerable Random Field (SRF), where the potentials are adapted to the local image structure. We build on the idea of defining MRF clique potentials in terms of linear filter responses [18, 29], which allows us to connect MRFs with the literature on steerable filters [3]. Instead of using a fixed set of filters, such as derivative filters in the pairwise case, or some other fixed set of filters in the high-order case [18, 29], we *steer* the

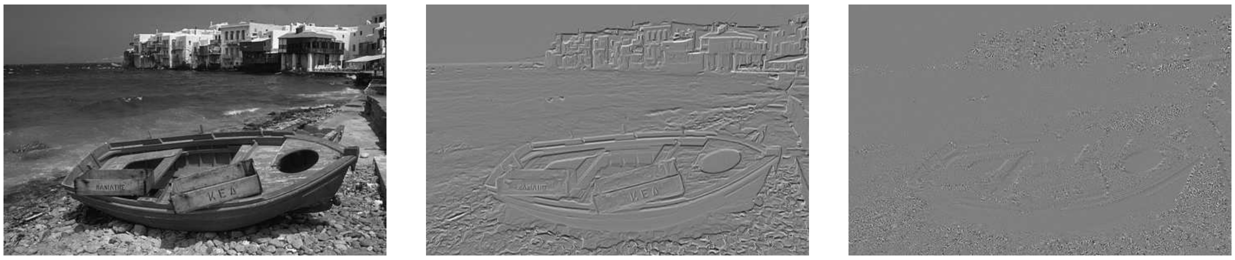


Figure 2. **Example image and steered derivatives.** The derivative response orthogonal to the local structure is shown in the middle, the response aligned with the image structure on the right.

filters to the local structure of the input image. Specifically, we steer horizontal and vertical image derivatives to the predominant orientation of the image structure, and obtain derivatives that are orthogonal to, and aligned with, the orientation[1]. Image derivatives are computed using extended (2×3 and 3×2) filters. Consequently, the SRF is a high-order random field based on extended cliques (*cf*. [18, 29]).

The local orientation is computed using the structure tensor approach [8], which has a long history in continuous-space anisotropic regularization methods (see, *e.g*., [27]). In this paper we combine and connect ideas from MRFs, steerable filters, and anisotropic diffusion. Specifically, we introduce the idea of steering into the domain of spatially discrete random field models. Doing so has several advantages: First, we can use the empirical statistics of the steered filter responses to motivate particular models. And second, we can learn SRF models from training data, which gives steerable (or anisotropic) regularization a statistical foundation, and allows us to avoid parameter tuning by hand.

Additionally we study the marginal statistics of steered filter responses (Fig. 1) and find that while both are heavy-tailed, the derivative orthogonal to the image structure has a much broader histogram than the aligned derivative. The SRF potentials model these steered filter responses using a Gaussian scale mixture (GSM) [25], which is able to capture their heavy-tailed characteristics.

We illustrate the SRF model in image restoration applications, but note that, as with any MRF, it applies much more generally, such as to stereo and image motion. In particular, we show that SRFs substantially outperform pairwise MRF models in image denoising and inpainting. One particular advantage of SRF models is that, by focusing the model on oriented image structure, they lead to a better restoration of image edges.

2. Previous Work

While steered derivatives [3] have been used in a variety of contexts including image restoration [27] and image coding [28], the statistics of steered derivative responses in generic scenes are not widely studied. Sidenbladh and Black [21] explored the "object specific" statistics of steered filter responses at locations in the image corresponding to the limbs of people. In generic scenes, Konishi *et al*. [9] studied the statistics of gradient responses and eigenvalues on, and off, image edges. Finally, Scharr *et al*. [20] modeled the statistics of the eigenvalues of the structure tensor in natural scenes and used these marginal statistics in a diffusion denoising method. In a comprehensive study of image contours, Elder and Goldenberg [2] considered many image properties related to edge continuity, but not specifically steered filter responses.

The proposed SRF is quite closely related to anisotropic [27] as well as classical edge-preserving [1] regularization techniques. In anisotropic diffusion, smoothing depends on the local orientation of the image structure. While the orientation is computed using the structure tensor, as in the SRF model, the amount of smoothing depends on the eigenvalues of the structure tensor or on the gradient magnitude, whereas in the SRF the amount of smoothing depends on steered image derivatives. [20] shows that these eigenvalues are smoothed versions of the (squared) steered derivatives. This is an important difference, because an energy function based on eigenvalues introduces a smoothing term in the corresponding diffusion algorithm, which leads to inferior results [20]. Moreover, the model from [20] is based on learning marginals of the eigenvalues, while the SRF uses full discriminative learning. Standard anisotropic diffusion is motivated in *algorithmic* terms and not through an energy function, and cannot directly be interpreted as being an approximate inference technique for a probabilistic model. Steerable Random Fields, on the other hand, are derived as a generic *probabilistic model* of images (or some other dense scene representation), which admits a wide range of uses beyond diffusion-based smoothing.

Related to SRFs is the work by Lyu and Simoncelli [14], which performs image denoising by modeling oriented wavelet responses using an MRF. In contrast to what is proposed here, they use a fixed set of orientations and do not explicitly adapt the potentials to the local structure orientation. Hammond and Simoncelli [5] model wavelet re-

[1]This steering can be based on the output of the model (images, depth maps, flow fields, *etc*.) or on the input. SRFs that are steered to the input are conditional models (CRFs) that directly model the posterior distribution of the output given the input image.

sponses in an orientation adaptive way, but do not propose a spatial model such as an MRF. Another aspect that relates both these works to ours is their use of Gaussian scale mixtures to capture heavy-tailed image statistics.

Levin *et al*. [12] perform image inpainting in the gradient domain, which allows them to regularize the image by penalizing changes in the gradient magnitude and the gradient orientation. In contrast to the SRF, they formulate a pairwise MRF in this gradient magnitude and orientation domain. For inpainting applications they learn the model from the local image statistics, whereas we define a generic model trained on a large database.

3. Steered Filter Responses and Their Statistics

The statistics of horizontal and vertical image derivatives in natural scenes have been extensively studied (see, *e.g*., [22]). These derivatives are found to have heavy-tailed marginal statistics that arise from the fact that, while neighboring pixels often have similar intensities, significant discontinuities arise quite frequently. Heavy-tailed distributions are characterized by large values of the kurtosis $\kappa = E[(x-\mu)^4]/E[(x-\mu)^2]^2$. Fig. 1(a),(b) shows marginal log-histograms of horizontal and vertical image derivatives. The histograms were computed from a set of 25 natural images taken from the Berkeley segmentation dataset [15] (only the luminance was considered). Simple first order derivative filters ($(1,-1)$ and $(1,-1)^{\mathrm{T}}$) were used. The histograms have much tighter peaks and heavier tails than a Gaussian with the same mean and variance.

To obtain the steered derivatives, we first need to estimate the orientation of the local image structure. For this we rely on the structure tensor [8]:

$$S = G_\rho * \nabla \mathbf{I} \cdot \nabla \mathbf{I}^{\mathrm{T}} = G_\rho * \begin{bmatrix} (\partial_x^{5\times 5}\mathbf{I})^2 & \partial_x^{5\times 5}\mathbf{I} \cdot \partial_y^{5\times 5}\mathbf{I} \\ \partial_x^{5\times 5}\mathbf{I} \cdot \partial_y^{5\times 5}\mathbf{I} & (\partial_y^{5\times 5}\mathbf{I})^2 \end{bmatrix}. \quad (1)$$

Here $\partial_x^{5\times 5}\mathbf{I}$ denotes the horizontal derivative of image $\mathbf{I}$ computed with the filter $D_x^{5\times 5}$ (see below), $\partial_y^{5\times 5}$ the vertical derivative, and G_ρ a Gaussian smoothing kernel with standard deviation ρ. To obtain reliable estimates of the local image orientation, we use optimized 5×5 derivative filters [19]:

$$\begin{aligned} D_x^{5\times 5} =& (0.0234, 0.2415, 0.4700, 0.2415, 0.0234)^{\mathrm{T}} \cdot \quad (2) \\ & (0.0838, 0.3323, 0, -0.3323, -0.0838) \\ D_y^{5\times 5} =& (D_x^{5\times 5})^{\mathrm{T}}. \quad (3) \end{aligned}$$

After estimating the structure tensor, we perform an eigendecomposition and obtain the eigenvectors $(\cos\theta, \sin\theta)^{\mathrm{T}}$ orthogonal to, and $(-\sin\theta, \cos\theta)^{\mathrm{T}}$ aligned with the local orientation. If a strong edge is present, the local orientation is the orientation of the edge. The derivative orthogonal to the local orientation, ∂_O, and the aligned derivative, ∂_A, are now given as

$$\partial_O = \cos\theta \cdot \partial_x + \sin\theta \cdot \partial_y \quad (4)$$

$$\partial_A = -\sin\theta \cdot \partial_x + \cos\theta \cdot \partial_y. \quad (5)$$

Here, ∂_x and ∂_y are the horizontal and vertical image derivatives, which are computed using 2×3 and 3×2 filters (see below). Fig. 2 shows an example image along with its steered derivatives. We can see that the derivative orthogonal to the local orientation shows continuous edge-like structures, while the aligned derivative has much less spatial structure.

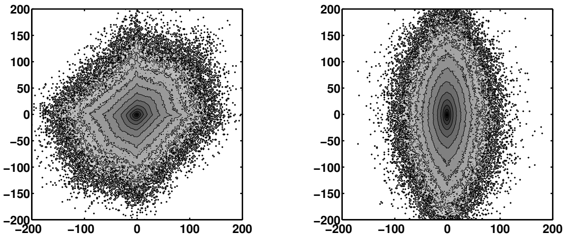


Figure 3. **Joint filter response log-histograms.** *(left)* Standard derivatives. *(right)* Steered derivatives.

The empirical steered derivative histograms are shown in Fig. 1(c),(d). The log-histogram of the derivative orthogonal to the local (*e.g*., edge) orientation has much broader tails than the aligned derivative; this indicates that, as expected, large intensity changes occur more frequently orthogonal to the edge orientation. While both log-histograms are heavy-tailed, the kurtosis of the aligned derivative is much higher.

These findings lead to two observations. First, the steered marginal statistics provide a statistical motivation for anisotropic diffusion methods. Standard anisotropic diffusion performs relatively strong linear smoothing along edges, and nonlinear smoothing that allows for large intensity jumps orthogonal to the edges. The tight marginal histogram of the aligned derivative suggests that we should indeed perform stronger smoothing aligned with edges than orthogonal to them. Second, and in contrast to what is done in standard anisotropic diffusion, the non-Gaussian character of the histograms indicates that we should employ nonlinear smoothing even aligned with the edges.

Finally, we can compare the joint log-histogram of steered derivatives in Fig. 3(b) to that of standard derivatives in Fig. 3(a). The log-histogram of horizontal and vertical derivatives exhibits polygonal isocontours as previously noted in the study of natural images [22], but the joint log-histogram of the steered responses has much more elliptical contour lines.

4. Steerable Random Field Model

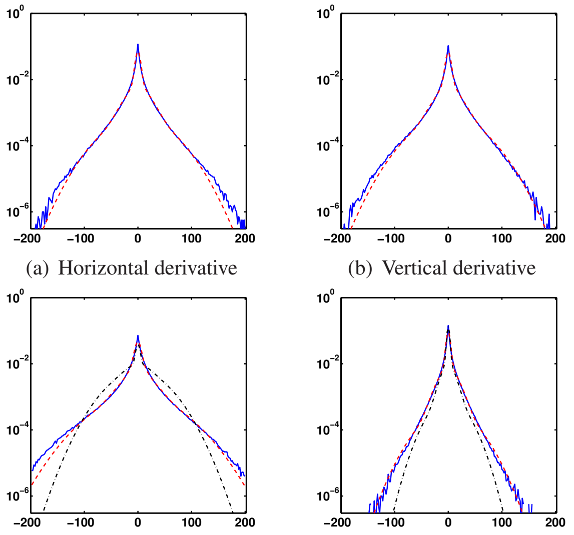
In the Steerable Random Field, we exploit statistical properties of steered derivatives to define a random field model of natural images. The model can be trained in two different ways, which we will discuss below: (1) by fitting the marginal distributions of the steered derivatives to model a prior over images; (2) using a discriminative variant of contrastive divergence [7] to model a posterior. Suppose that our goal is to recover an artifact-free, true, or "hidden" image $\mathbf{H}$ from a corrupted or "noisy" observation, the input $\mathbf{I}$. In a Bayesian approach we typically represent the posterior distribution as $p(\mathbf{H}|\mathbf{I}) \propto p(\mathbf{I}|\mathbf{H}) \cdot p(\mathbf{H})$, where the likelihood $p(\mathbf{I}|\mathbf{H})$ describes the observation process that took the true image $\mathbf{H}$ and corrupted it to obtain the observation image $\mathbf{I}$. The prior $p(\mathbf{H})$ models the a-priori probability density of having a particular underlying true image $\mathbf{H}$ among all possible images. Very often, such priors are formulated as Markov random fields.

We take a slightly different approach here and model the posterior distribution directly, but still break it into two components:

$$\begin{aligned} p(\mathbf{H}|\mathbf{I};\Omega) &= \tfrac{1}{Z(\mathbf{I})} \exp\{-E(\mathbf{H},\mathbf{I};\Omega)\} \\ &= \tfrac{1}{Z(\mathbf{I})} f_L(\mathbf{I};\mathbf{H}) \cdot f_S(\mathbf{H};\mathbf{I},\Omega). \end{aligned} \tag{6}$$

Here, $f_L(\mathbf{I};\mathbf{H})$ is a (unnormalized) distribution that models the observation process, $f_S(\mathbf{H};\mathbf{I},\Omega)$ is a (unnormalized) steerable image model, Ω are the model parameters, and $Z(\mathbf{I})$ is the so-called partition function that ensures that the posterior is normalized. Since, in applying the model, we will attempt to maximize the posterior w.r.t. $\mathbf{H}$, we do not require the partition function to be known during inference, and hence ignore it for the most part. For now we simply assume that we have a suitable observation model for the task we are trying to solve; we will return to details on the observation model in Section 5. One important point to note here is that in this architecture neither the observation model nor the image model by themselves are good models of the posterior distribution. Only their combination results in a good Bayesian model, *e.g.*, for image restoration.

Basic model. The steerable image model $f_S(\mathbf{H};\mathbf{I},\Omega)$ intuitively takes the role of the prior from above in that it assesses how "natural" the true image $\mathbf{H}$ is. In contrast to the standard approach, the image model is not truly a prior distribution here, because it depends on the image observation $\mathbf{I}$. The SRF thus has a conditional random field (CRF) architecture (*cf*. [6, 10, 11, 23]). As in a standard MRF-based formulation, we assume that the pixels are described by nodes in a graph. Usually, the neighborhood structure of the edges describes the factorization structure of the image model. To simplify the discussion, we make the factorization more explicit, and instead describe the model as a factor



(a) Horizontal derivative (b) Vertical derivative

(c) Orthogonal derivative (d) Aligned derivative

Figure 4. **Fit of GSM model** with 4 scales (dashed, red) to filter response statistics (solid, blue). The GSM model learned with contrastive divergence is shown in case of the steerable model (dash-dotted, black).

graph [24]. We assume that the true image $\mathbf{H}$ is subdivided into overlapping patches of $m \times n$ pixels; $\mathbf{H}_{(j)}$ denotes the vector of all pixels in the j^{th} patch. We write the steerable image model as a product over all patches in the image

$$f_S(\mathbf{H};\mathbf{I},\Omega) = \prod_j f_{S,j}(\mathbf{H}_{(j)};\mathbf{I},\Omega). \tag{7}$$

If we disregard the dependence on $\mathbf{I}$ then each factor corresponds to a clique in a classical MRF prior of images. But unlike a standard pairwise prior, which models each factor using either horizontal or vertical image derivatives, and unlike the Field of Experts (FoE) [18], which models each factor using filter responses from a set of fixed filters, the SRF models each factor using responses from adaptive, steered filters. For that, let $\theta_j = \theta_j(\mathbf{I})$ denote the orientation of the image structure at position j (*i.e.*, the center of the patch $\mathbf{H}_{(j)}$) as determined from the eigendecomposition of the structure tensor of $\mathbf{I}$. Then we write each factor of the SRF model as

$$\begin{aligned} f_{S,j}(\mathbf{H}_{(j)};\mathbf{I},\Omega) =& \phi_O(\cos\theta_j \cdot \partial_x\mathbf{H} + \sin\theta_j \cdot \partial_y\mathbf{H};\Omega) \cdot \\ & \phi_A(-\sin\theta_j \cdot \partial_x\mathbf{H} + \cos\theta_j \cdot \partial_y\mathbf{H};\Omega), \end{aligned}$$

where $\phi_O(\cdot)$ is a model of the derivative response orthogonal to the structure orientation, and $\phi_A(\cdot)$ models the derivative aligned with the structure. $\partial_x\mathbf{H}$ and $\partial_y\mathbf{H}$ denote the horizontal and vertical image derivatives, which are evaluated at position j and computed from the image patch $\mathbf{H}_{(j)}$.

In order to model the potentials, we rely on a Gaussian scale mixture (GSM) representation [25], which has been used in various models of natural images, *e.g.* [14]. One advantage of GSM models is that they are computationally relatively easy to deal with, yet flexible enough to represent heavy-tailed distributions such as the ones shown in Fig. 1. In particular, we write the model for the orthogonal derivative as

$$\phi_O(r;\Omega) = \sum_k \omega_{O,k} \cdot \mathcal{N}\left(r;\ 0, \sigma_O^2/s_k\right). \qquad (8)$$

In this formulation, $\omega_{O,k}$ is the mixture weight for the k^{th} Gaussian component with mean 0, base variance σ_O^2, and scale s_k. The model for the aligned derivative is equivalent (but using $\omega_{A,k}$ and σ_A). We assume that the base variance and scales are fixed and only regard $\Omega = (\omega_{O,k}, \omega_{A,k} | k = 1, \ldots, K)$ as the set of parameters of the GSM model for the steered derivatives.

Both as a sanity check, and as a simple way of learning SRF models from data, we fit GSM models to the empirical marginals. Fig. 4 shows how both the standard and the steered marginal derivative statistics are fit well by GSM models with 4 scales. In each case the base variance σ_O^2 was chosen to be the variance of the respective empirical marginal, and scales $s = (0.008, 0.04, 0.1, 1)^{\mathrm{T}}$ were used. The mixture weights were fit using a simple expectation maximization (EM) procedure.

Practical considerations. We may now be tempted to define the model using factors on overlapping 3×3 patches, and to use standard derivative filters of the type $(1/2, 0, -1/2)$ (with possible smoothing in the orthogonal direction). Doing so, however, leads to checkerboard-like artifacts in image restoration, as has been observed in [20]. The problem is that standard 3-tap derivative filters effectively decouple directly neighboring pixels (because of the 0 center coefficient) leading to artifacts. This could be corrected using an additional smoothing term as suggested in [20], but we would like to avoid that to be able to recover crisp boundary structures. We could use $(1, -1)$ type derivative filters to avoid this problem, but then the x- and y-derivatives are not computed at the same spatial location.

To solve this problem we define *two* kinds of filters, one pair of 2×3 derivative filters, and one pair of 3×2 derivative filters:

$$D_x^{2\times3} = \tfrac{1}{2}\begin{pmatrix} 1 & 0 & -1 \\ 1 & 0 & -1 \end{pmatrix} \qquad D_y^{2\times3} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & -1 & 0 \end{pmatrix} \qquad (9)$$

$$D_x^{3\times2} = (D_y^{2\times3})^{\mathrm{T}} \qquad D_y^{3\times2} = (D_x^{2\times3})^{\mathrm{T}}. \qquad (10)$$

These filters have the advantage that they do not have a 0 center coefficient in one of the directions, but each pair estimates x- and y- derivatives at the same spatial location. We can combine these filters in the SRF framework by having *two* kinds of factors with their corresponding image patches: One consists of all overlapping patches of 2×3 pixels; the other one of all overlapping 3×2 patches. Depending on which patch-size $\mathbf{H}_{(j)}$ has, the factor $f_{S,j}(\mathbf{H}_{(j)}; \mathbf{I})$ uses the appropriate derivative filters. Fig. 5 shows the corresponding factor graph with both types of factors.

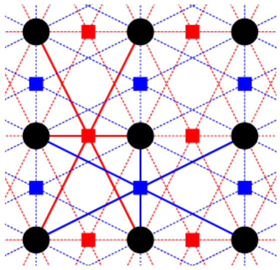


Figure 5. **Factor graph representation of the SRF model** (only nodes of $\mathbf{H}$ shown). The red factor nodes and edges indicate the 3×2 factors; the blue ones the 2×3 factors. One factor each is highlighted with solid bold edges.

Learning. To estimate the model parameters (*i.e.*, the mixture weights) from data, we could either simply fit the empirical marginals of the steered derivatives as above, or we could use a more rigorous learning procedure, such as maximum likelihood. In case of a pairwise MRF, fitting marginals has a rigorous interpretation, because, as Wainwright and Jordan [24] showed, marginals of a pairwise model are fixed points of the Bethe free energy (we reasonably assume here that the intensity distribution of each pixel is uniform). Hence we train the pairwise models used for the comparisons in the next section in this way. The SRF, on the other hand, is a high-order RF model, where this interpretation can not easily be exploited.

Because of the conditional nature of the SRF model, we train it in the context of a particular application represented by a suitable likelihood model. More precisely, we try to maximize the conditional log-likelihood of the training data $H = \{\mathbf{H}^{(1)}, \ldots, \mathbf{H}^{(t)}\}$ given the "noisy" observation data $I = \{\mathbf{I}^{(1)}, \ldots, \mathbf{I}^{(t)}\}$ with respect to the model parameters (mixture weights) Ω:

$$\mathcal{L}(H|I;\Omega) = \sum_l \log p(\mathbf{H}^{(l)} | \mathbf{I}^{(l)}; \Omega). \qquad (11)$$

Unfortunately maximizing the (conditional) likelihood in loopy graphical models is difficult since the partition function cannot easily be computed. Learning thus needs to rely on approximations, often using Markov chain Monte Carlo (MCMC) methods [29], which tend to be very slow. An alternative is the contrastive divergence (CD) algorithm [7], which has already been applied to a number of random field models in vision [6, 18]. We employ a discriminative variant for training CRF models (see [6] for details). The advantage of CD over (conditional) maximum likelihood is that the Markov chain does not have to be run until convergence, but only for a small number of steps. As suggested in [18], we use a hybrid Monte Carlo sampler with 30 leaps, $l = 1$ MCMC steps, and a learning rate of 0.01. We ran CD for 2000 iterations after which the mixture parameters did not change significantly. The derivation of the gradient of the posterior energy w.r.t. the mixture weights is relatively straightforward and is omitted due to space constraints.

Non-conditional model. While we do not explore this formulation rigorously here, it is straightforward to define the SRF model as a pure prior model without conditioning on the observation. In this case, steering is done using the true, underlying image, and training is done in a fully generative way, for example using contrastive divergence.

5. Applications

To compare the Steerable Random Field model with traditional pairwise MRF models, we apply them to two different image restoration tasks, image denoising and inpainting.

For denoising, we assume that the image observation $\mathbf{I}$ is corrupted with additive $i.i.d.$ Gaussian noise, as is standard for denoising performance evaluations. We thus formulate the observation model as $f_L(\mathbf{I}; \mathbf{H}) = \prod_j \mathcal{N}(H_j;\ I_j, \sigma_L^2)$. For our experiments, we assume a fixed noise standard deviation of $\sigma_L = 20$. We obtain a full posterior model of denoised images (*cf*. Eq. (6)) by combining this observation model with the SRF model ($\rho = 1$) from Eq. (7). We then train the parameters of the SRF model discriminatively using contrastive divergence.

The training data consists of a set of "clean" training images, H, and a set of noisy training images, I, which have been obtained from H by adding Gaussian noise drawn from the observation model. The clean image data consists of 20000 image patches of 9×9 pixels, which have been randomly cropped from the 25 natural images used in Section 3. Note that the structure tensor was computed before the images were cropped to avoid boundary artifacts.

Fig. 4(c),(d) shows the GSM response models determined by this procedure. Even though the shape of the learned model differs quite a bit from the marginal distribution, it performs much better than a GSM model fit directly to the marginals[2]. For comparison with a baseline technique, we use a traditional pairwise MRF model, where we formulate the potentials using Gaussian scale mixtures as in Eq. (8). To clearly separate the contribution of the steerable filters versus the high-order cliques, we also used a high-order, non-steerable MRF based on 2×3 and 3×2 derivative filters (referred to as "Non-steerable MRF"). In both cases, we use the EM algorithm and the scale parameters described in Section 4.

To perform inference we rely on simple continuous local optimization methods for both the SRF and the baseline models. While inference in pairwise MRFs can be done using belief propagation or graph cuts, doing so for high-order RF models, such as the SRF, is considerably more

[2]Zhu *et al*. [29] also observed that the optimal clique potential functions do not necessarily correspond to the marginal statistics.

(a) Image denoising $\sigma = 20$

Model / Algorithm	PSNR in dB	SSIM
Pairwise MRF	27.726	0.7672
Non-steerable MRF	27.707	0.7703
SRF	28.316	0.7878
SRF with ST update	28.299	0.7926

(b) Image inpainting

Model / Algorithm	PSNR in dB	SSIM
Pairwise MRF	30.786	0.940
SRF	31.578	0.944
SRF with ST update	32.208	0.949

Table 1. **Average performance on 68 test images.**

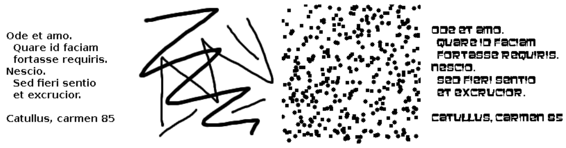
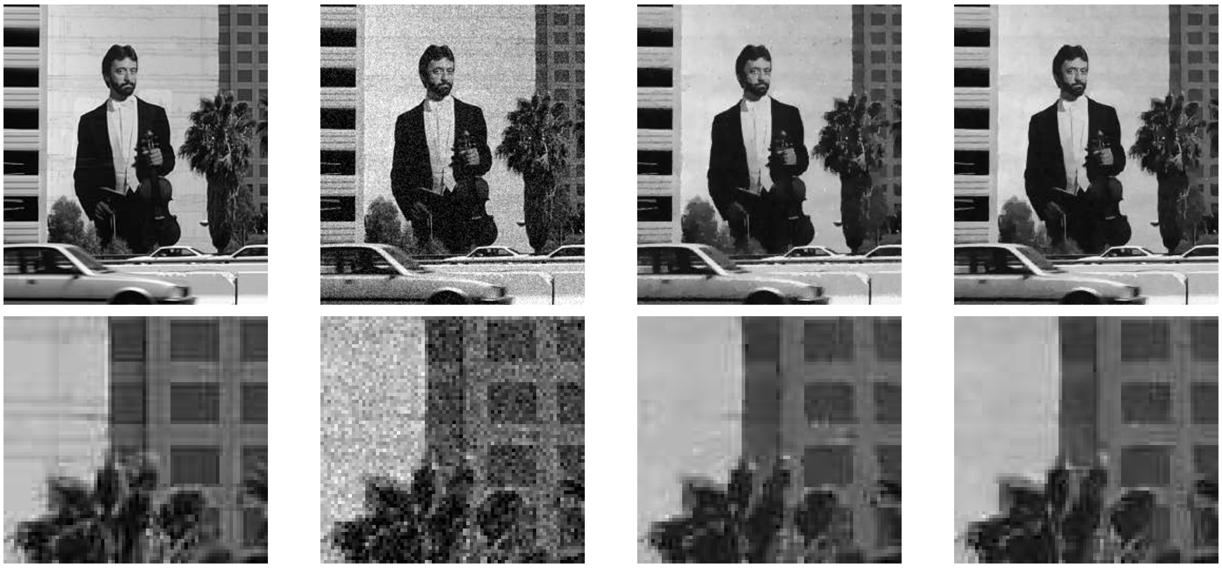


Figure 7. **Inpainting masks used to create test set.**

difficult [16]. To simplify matters, we use the posterior energy $E(\mathbf{H}, \mathbf{I}; \Omega)$ and maximize it w.r.t. $\mathbf{H}$ using a conjugate gradient method [17]. The gradient expressions are relatively straightforward to derive and omitted here for brevity. As has been noted before (*e.g.*, [18]), denoising quality can typically be improved by appropriately weighting the relative strength of the observation model versus the regularization term. In that vein, we use $f_S(\mathbf{H}; \mathbf{I}, \Omega)^\alpha$ as the regularization term for denoising (similarly for the pairwise and non-steerable MRF); the optimal α value is determined by denoising 10 training images with a set of candidate values and choosing the best one. To determine the best α, and also to measure denoising performance on the test set, we use the peak signal-to-noise ratio (PSNR). In addition, we also measure test performance using the perceptually-based structural similarity index (SSIM) [26].

Table 1 shows denoising results averaged over the same 68 test images used in [18]. The SRF substantially outperforms both the traditional pairwise and the high-order, non-steerable MRF model, each in terms of PSNR and SSIM, despite using the same number of filters. This clearly suggests that the performance improvement from the SRF is due to the use of steerable filters. We also tried updating the structure tensor using the estimate of the noise free image during inference ("ST update"), which results in no significant difference in performance. Fig. 6 shows denoising results for one of the test images. The detail images reveal that the SRF is much better at recovering continuous edge structures while also preserving more detail (*e.g.*, the lines on the wall in the background). It is worth noting that, due to the particular derivative filters, the results do not exhibit checkerboard-like artifacts that have been a problem with certain anisotropic diffusion schemes (*e.g.*, [20]).



(a) (b) (c) (d)

Figure 6. **Image denoising using a SRF.** (a) Original image. (b) Gaussian noise added ($\sigma = 20$, PSNR = 22.49dB, SSIM = 0.528). (c) Denoised with a pairwise MRF (PSNR = 27.60dB, SSIM = 0.810). (d) Denoised with a SRF (PSNR = 28.63dB, SSIM = 0.836, without "ST update"). The bottom row shows a cropped version of the image above.

One disadvantage of discriminatively trained models is that they need to be trained end-to-end for each application, potentially even for each noise level, although compared to generative models, they focus their modeling power on what matters the most for the application at hand. To show that the SRF model trained on one particular application nevertheless generalizes to another application, we apply the model with the parameters obtained by training on the image denoising task to an image inpainting application. We assume that the user has specified a mask $\mathcal{M}$ of pixels to be filled in, and we use the same likelihood model used in [18]. Since we compute the structure tensor from the masked input image, we use a larger structure tensor with $\rho = 4$, which is sufficient to estimate the orientation in the areas of missing pixels. As for denoising, we use conjugate gradient optimization for inference. To evaluate the performance of the SRF, we generate a synthetic set of inpainting tasks by masking pixels from the same 68 test images used above. Fig. 7 shows the 4 different masks used to generate the test set. Table 1 summarizes the average performance on all test images. The SRF model performs substantially better than a pairwise MRF model. In this case, the performance can be further improved by steering the structure tensor based on the inpainted image ("ST update"). Fig. 8 shows one of the test images, which illustrates how the steered derivatives help smoothly complete image contours.

6. Summary and Future Work

In this paper we introduced the idea of using steerable filters into the domain of spatially discrete random field models of images and other dense scene representations. In contrast to previous MRF models, which formulate clique potentials based on fixed sets of filters, the proposed Steerable Random Field uses steered filter responses computed using a structure tensor. Based on a study of the statistics of steered filter responses, we formulated the potentials of this SRF model using Gaussian scale mixtures, whose parameters were learned from training data with a discriminative version of contrastive divergence. The proposed model thus unifies various ideas from RF modeling and anisotropic regularization with structure tensors, and puts the latter on a statistical foundation.

We demonstrated how the SRF can be used in image restoration applications with substantial performance gains over pairwise MRFs and other non-steered models. In particular, edge-like structures are restored much more cleanly using the SRF. We should note that the denoising results are about 0.4dB worse than those reported with Fields of Experts [18]. This is not a surprise given the fact that SRFs use only 2 filters and smaller cliques; nor is this a major goal of this work. The main goal here is to introduce spatially-adaptive steering to MRF models; future work will have to consider whether more complex models with many filters, such as the FoE, can be steered as well.

By connecting algorithmic anisotropic diffusion techniques with probabilistic MRF models, the SRF opens the possibility of employing new learning and inference methods, such as belief propagation [16], in these problems. Future work should consider applying such techniques to SRFs.

(a) (b) (c) (d)

Figure 8. **Image inpainting using a SRF.** (a) Masked image (red regions are to be filled in). (b) Inpainting with a pairwise MRF (PSNR = 37.83dB, SSIM = 0.980). (c) Inpainting with a SRF (PSNR = 41.08dB, SSIM = 0.983, without "ST update"). (d) Detail results. Top row: Original image, masked image. Bottom row: Inpainting with pairwise MRF, inpainting with SRF.

Acknowledgments This work was supported by NSF grants IIS-0534858 and IIS-0535075. We would like to thank Siwei Lyu for interesting discussions.

References

[1] P. Charbonnier, L. Blanc-Feéraud, G. Aubert, and M. Barlaud. Deterministic edge-preserving regularization in computed imaging. *IEEE T. Image Process.*, 6(2):298–311, 1997.

[2] J. H. Elder and M. Goldenberg. Ecological statistics of Gestalt laws for the perceptual organization of contours. *J. of Vision*, 2:324–353, 2002.

[3] W. T. Freeman and E. H. Adelson. The design and use of steerable filters. *PAMI*, 13(9):891–906, 1991.

[4] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *PAMI*, 6:721–741, 1984.

[5] D. K. Hammond and E. P. Simoncelli. Image denoising with an orientation-adaptive Gaussian scale mixture model. *ICIP 2006*, pp. 1433–1436.

[6] X. He, R. S. Zemel, and M. A. Carreira-Perpiñán. Multiscale conditional random fields for image labeling. *CVPR 2004*, v. 2, pp. 695–702.

[7] G. E. Hinton. Training products of experts by minimizing contrastive divergence. *Neural Comput.*, 14(8):1771–1800, 2002.

[8] H. E. Knutsson, R. Wilson, and G. H. Granlund. Anisotropic nonstationary image estimation and its applications: Part I–Restoration of noisy images. *IEEE T. Comm.*, 31(3):388–397, 1983.

[9] S. M. Konishi, A. L. Yuille, J. M. Coughlan, and S. C. Zhu. Fundamental bounds on edge detection: A information theoretic evaluation of different edge cues. *CVPR 1999*, pp. 573–579.

[10] S. Kumar and M. Hebert. Discriminative random fields. *IJCV*, 68(2):179–201, 2006.

[11] J. Lafferty, A. McCallum, and F. Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. *ICML 2001*, pp. 282–289.

[12] A. Levin, A. Zomet, and Y. Weiss. Learning how to inpaint from global image statistics. *ICCV 2003*, v. 1, pp. 305–312.

[13] S. Z. Li. *Markov Random Field Modeling in Image Analysis*. Springer, 2nd edition, 2001.

[14] S. Lyu and E. P. Simoncelli. Statistical modeling of images with fields of Gaussian scale mixtures. *NIPS*2006*.

[15] D. Martin, C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. *ICCV 2001*, v. 2, pp. 416–423.

[16] B. Potetz. Efficient belief propagation for vision using linear constraint nodes. *CVPR 2007*.

[17] C. E. Rasmussen. `minimize.m`. `http://www.kyb.tuebingen.mpg.de/bs/people/carl/code/minimize/`.

[18] S. Roth and M. J. Black. Fields of experts: A framework for learning image priors. *CVPR 2005*, v. 2, pp. 860–867.

[19] H. Scharr. Optimal filters for extended optical flow. *First International Workshop on Complex Motion*, v. 3417 of *LNCS*. Springer, 2004.

[20] H. Scharr, M. J. Black, and H. W. Haussecker. Image statistics and anisotropic diffusion. *ICCV 2003*, v. 2, pp. 840–847.

[21] H. Sidenbladh and M. J. Black. Learning the statistics of people in images and video. *IJCV*, 54(1–3):183–209, 2003.

[22] A. Srivastava, A. B. Lee, E. P. Simoncelli, and S.-C. Zhu. On advances in statistical modeling of natural images. *J. Math. Imaging Vision*, 18(1):17–33, 2003.

[23] B. Taskar, C. Guestrin, and D. Koller. Max-margin Markov networks. *NIPS*2003*.

[24] M. J. Wainwright and M. I. Jordan. Graphical models, exponential families, and variational inference. Technical Report 649, Dept. of Statistics, UC Berkeley, 2003.

[25] M. J. Wainwright and E. P. Simoncelli. Scale mixtures of Gaussians and the statistics of natural images. *NIPS*1999*, pp. 855–861.

[26] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE T. Image Process.*, 13(4):600–612, 2004.

[27] J. Weickert. *Anisotropic Diffusion in Image Processing*. Teubner Verlag, Stuttgart, Germany, 1998.

[28] R. Wilson, H. E. Knutsson, and G. H. Granlund. Anisotropic nonstationary image estimation and its applications: Part II–Predictive image coding. *IEEE T. Comm.*, 31(3):398–406, 1983.

[29] S. C. Zhu, Y. Wu, and D. Mumford. Filters, random fields and maximum entropy (FRAME): Towards a unified theory for texture modeling. *IJCV*, 27(2):107–126, 1998.