

Sparse Long-Range Random Field and Its Application to Image Denoising

Yunpeng Li and Daniel P. Huttenlocher

Department of Computer Science, Cornell University, Ithaca, NY 14853
{yuli, dph}@cs.cornell.edu

Abstract. Many recent techniques for low-level vision problems such as image denoising are formulated in terms of Markov random field (MRF) or conditional random field (CRF) models. Nonetheless, the role of the underlying graph structure is still not well understood. On the one hand there are pairwise structures where each node is connected to its local neighbors. These models tend to allow for fast algorithms but do not capture important higher-order statistics of natural scenes. On the other hand there are more powerful models such as Field of Experts (FoE) that consider joint distributions over larger cliques in order to capture image statistics but are computationally challenging. In this paper we consider a graph structure with longer range connections that is designed to both capture important image statistics and be computationally efficient. This structure incorporates long-range connections in a manner that limits the cliques to size 3, thereby capturing important second-order image statistics while still allowing efficient optimization due to the small clique size. We evaluate our approach by testing the models on widely used datasets. The results show that our method is comparable to the current state-of-the-art in terms of PSNR, is better at preserving fine-scale detail and producing natural-looking output, and is more than an order of magnitude faster.

1 Introduction

Random fields are among the most common models used in low-level vision problems such as image restoration, segmentation, and stereo. The strength of these models lies in their ability to represent both the interaction between neighboring pixels and the relationship between the observed data values and estimated labels at each pixel. A random field model defines a graph structure with potential functions over the labelings of cliques in this graph. For low-level vision problems the graph generally has a node corresponding to each pixel, edges connecting certain pairs of neighboring pixels, and potentials that encourage neighboring pixels to have similar labels. Two key issues in the application of random field models to a given problem are (i) defining appropriate graph structures and (ii) finding suitable potential functions over the cliques of that graph. Most research has focused on the latter problem, whereas here we focus on the former.

In this paper we study *sparse long-range random field* (SLRF) models, that represent interactions between distant pixels using sparse edges so as to maintain a fixed clique size. The size of the clique is chosen so as to be appropriate for a particular problem. In image denoising, second-order spatial terms are important for representing intensity

change. Thus we use a graph structure that has cliques of size three, as discrete approximations to a second order function require three data values. In this framework the potential functions are defined over fixed-size cliques that have different spatial extents, effectively encoding image structure of a fixed order (defined by the clique size) at multiple scales of observation. This enables such models to produce smooth labelings for highly noisy images but at the same time allows efficient solution.

In contrast, other recent work using higher-order models and longer-range connections, such as the Field of Experts (FoE) model [1], has large cliques and thus does not support fast optimization. Our main interest is thus in investigating whether simpler models with smaller cliques can produce results comparable to the state-of-the-art achieved with more powerful models, such as FoE, while using much less time. The experiments that we report here, performed on widely used datasets, indicate that this is indeed the case. Not only do we achieve comparable peak signal-to-noise ratio (PSNR) to large-clique methods, our method is also better at avoiding over-smoothing although that is not captured by the PSNR measure. At the same time, our method is over 100 times faster than FoE and at least 10 times faster than other spatial-domain methods that achieve state-of-the-art results.

1.1 Motivation and Related Work

Random field models in statistics have existed for decades [2] and also have a long history in computer vision (e.g. [3,4]). A Bayesian formulation of pixel labeling problems using a Markov random field (MRF) model decomposes the problem into a *prior* that enforces spatial consistency of the labels, and a *likelihood* function that encourages agreement between the labels and the observed data. We use the more general terminology “spatial term”, rather than “prior”, and “data term”, rather than “likelihood”, as this applies to both Bayesian and non-Bayesian models, as well as to models that do not have a probabilistic interpretation.

In recent years, MRF models have been a popular choice for many low-level vision problems such as image restoration (e.g., [1,5,6]). The resurgence in the use of MRF models is largely complemented by the development of efficient approximation algorithms for inference on these models, including discrete methods such as loopy belief propagation (LBP) [7], tree-reweighted message passing [8,9] and graph cuts (GC) [10], as well as gradient-based methods such as diffusion [1] and variational inference [6]. Each of these methods has its own pros and cons, some of which have been investigated in [11]. The more recent conditional random field (CRF) [12] was originally proposed as a tree-structured model to address the label bias of hidden Markov models (HMM) in the context of natural language processing, and has since also been applied to loopy graphs. As a discriminative model, CRF is more convenient in situations where the generative process is unclear, the spatial term (i.e. the *prior*) depends on the observations, or the label of one site is related to observations at multiple sites (e.g., [13,14]).

The most widely used graph structure for random field models in low-level vision is a grid where each node is connected to its four immediate neighbors in the horizontal and vertical direction. While this model is simple and convenient to set up and optimize, it suffers from a number of drawbacks. First, as noted above, it can only represent first-order properties of the image, because it is based on pairwise relations between pixels.

Second, it can be sensitive to noise. Consider a connected region of n nodes in a 4-connected grid graph. In this case there are only approximately $O(\sqrt{n})$ connections between nodes in the region and those outside the region, because the boundary grows approximately as the square root of the area. Thus the data term of the n nodes over the region comes to dominate the connections outside the region, especially when robust (i.e. discontinuity preserving) spatial terms are used. For example, in image denoising this can be problematic for high noise levels because good estimates require substantial sized regions. Another way to view this is in terms of the standard deviation of the mean over the region. For concreteness, consider an image with additive Gaussian noise of $\sigma = 25$, and a 5×5 region of the image. The standard deviation of the mean of that region is $\sigma/\sqrt{5 \cdot 5} = 5$. At the same time, the perimeter-to-area ratio of such a neighborhood is only $4 \cdot 5/5^2 = 4/5$, or $1/5$ that of a single pixel. Hence the collective labeling of the group is dominated by its data term, which is subject to a non-trivial standard deviation of 5 in its mean.

The 4-connected grid graph is a special case of graphs that connect a node to all nodes that lie within some distance d . In contrast to our approach, such graphs produce quite dense edges even for moderate values of d . Early work on MRF models in vision, such as [15], used these models but restrict their attention to pairwise clique potentials. However, such pairwise models do not always capture the underlying distribution. For image denoising, in particular, second-order statistics are important, implying a need for cliques of size at least three.

Problems with earlier pairwise random field models have led to higher-order models such the Field of Experts (FoE) [1], where overlapping blocks of pixels are used rather than purely local connections. However, such models are computationally intensive due to their relatively large complete subgraphs. In addition, the learnt priors are also unintuitive, despite recent interpretations as derivative filters [6] and frequency filters [16]. This motivates our approach, which uses long-range connections to address the problem of noise but does so in the context of a simple graph structure with cliques of size three, so as to efficiently encode both first- and second-order spatial properties.

The work bearing the most similarity to ours is that of [17], which uses long-range edges in the problem of texture synthesis. Clique families are chosen using heuristic search based on the strength of interaction, which is evaluated on the training data. However, the model is restricted to pairwise clique potentials. Moreover each model is trained to synthesize a particular type of texture, which usually consists of some characteristic (and often repeating) patterns. Thus it is not well suited to modeling generic structures, such as those of natural scenes.

2 Sparse Long-Range Random Field

We now introduce our model. A sparse long-range random field (SLRF) is constructed so as to have a fixed clique size regardless of the spatial extent of the edges in the grid. Consider a set of nodes \mathcal{V} arranged on a grid, where there is a spatial distance defined between each pair of nodes. By choosing edges that increase in length exponentially, we can construct a graph that has a fixed clique size even though there is no bound on the maximum edge length. Consider the case of cliques of size 3, which as noted above (and

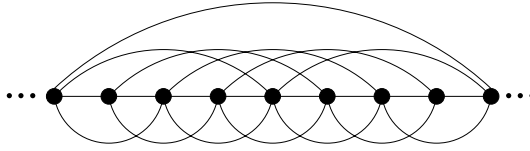


Fig. 1. Horizontal 3-cliques of \mathcal{E}_4^2 with edge lengths $\{1, 2, 4, 8\}$

discussed in more detail below) are important for image restoration because they enable the representation of second-order spatial statistics. A local 3-clique has edges of length 1 that connect each node to its immediate neighbors and edges of length 2 that connect each node to those two-away. Adding edges of length 4 to each node would then create additional 3-cliques composed of edges of length 2, 2 and 4, but does not increase the maximum clique size. Similarly for edges of length 8 and so on, as illustrated for the one-dimensional case in Figure 1.

More formally, each node is connected to other nodes at distance 2^k away from it, for integer values k such that $0 \leq k < K$. In other words the density of connections decreases exponentially with distance. We let \mathcal{E}_K^2 denote this set of edges; for instance, \mathcal{E}_4^2 is the set of edges of length $\{1, 2, 4, 8\}$. More generally, one could consider graphs where the edges are of length b^k for some $b > 2$ which yields sparser graphs. However, for $b = 3$ the resulting graphs already have maximum cliques of only size 2, which for image denoising does not allow representing second-order image statistics.

In the case of a two-dimensional image grid, edges may correspond to any underlying orientation. Considering both horizontal and vertical directions using edges in \mathcal{E}_K^2 again yields a graph with maximum cliques of size 3. These cliques correspond to spatial neighborhoods at different scales of observation and at different orientations (horizontal or vertical), but in each case capture second-order spatial information based on three nodes.

The inclusion of long-range edges in the SLRF offers the following advantages over a local grid model:

- *Improved information flow.* The graph requires fewer hops to reach one node from another, as is illustrated in Figure 1. In the example shown in the figure, the maximum graph distance between any two nodes is 2. Without long range edges, the corresponding numbers would be 8. In general, it can be shown that any two nodes v_1 and v_2 with grid distance d have graph distance $\mathcal{O}(d/b^{K-1} + bK)$. The decreased graph distance facilitates the flow of information over the random field.
- *Increased resistance to noise.* Long-range edges address the local bias problem discussed in Section 1. For any $n \times n$ neighborhood S with n up to b^{K-1} (i.e. up to the length of the longest edges), each node in S is connected to at least four nodes outside of S . Hence the total amount of interaction between S and the environment is now proportional to the area of S instead of its perimeter as in the 4-connected grid. This makes the strength of the spatial constraints between pixel blocks comparable to that of the data term over the block, suppressing noise-induced local bias without resorting to increasing the weight of the spatial term (which tends to cause over-smoothing).

The sparse nature of the SLRF also has the following computational benefits:

- *Small, fixed clique size.* As previously discussed, the size of the maximal cliques in an SLRF is either 2 or 3 regardless of the span of the longest range interaction being modeled. The low clique size allows arbitrary clique potentials to be optimized globally using efficient approximation algorithms such as belief propagation. In contrast, high-order random fields in general can only be optimized with continuous methods that rely on gradient (e.g. diffusion [1]), which may not exist in problems with discrete labels. Even when gradient-based methods are applicable, the running time is still super linear in the size of the cliques.¹
- *Low computational cost.* Since SLRF models have only K different edge lengths in an exponential series, the total number of edges in an SLRF is no more than K times of that in the underlying grid. Hence an SLRF model is at most $\log_b d$ times as costly as one with only short edges, where d is the length of the longest ranged interaction to be modeled. If on the other hand each node is connected to all the nodes near it up to some distance d (such as in [15]), the resulting graph would have $\Theta(d^2)$ edges and hence much higher computational. Although the model can still be called “sparse” from a graph theoretical point of view (as any graph with edge density independent of its size will qualify), it is clearly not so from the aspect of efficient optimization.

2.1 Cliques and Clique Potentials

Let $\mathcal{C} = \mathcal{C}_K^2$ denote the set of all cliques in an SLRF with edges \mathcal{E}_K^2 for a fixed K . There are several distinct types of cliques in this set, which can be characterized by the lengths of their edges. For instance,

$$\mathcal{C}_K^2 = C_{1,1,2} \cup C_{2,2,4} \cup \dots \cup C_{2^{K-2}, 2^{K-2}, 2^{K-1}} \quad (1)$$

where $C_{a,b,c}$ is the set of 3-cliques with edge length a , b , and c . Each of these sets of 3-cliques corresponds to observations at a different spatial scale, based on the lengths of the edges. Let $T(c)$ denote the type of clique c , e.g. $T(c) = (1, 1, 2) \forall c \in C_{1,1,2}$ and $T(c) = (1) \forall c \in C_1$.

We represent the likelihood of the random field as an exponential family of cost functions f and g parameterized by θ , where $f_{T(c)}^\theta$ is the spatial term and g^θ is the data term. Thus given observation I ,

$$p_\theta(X|I) = \frac{1}{Z(\theta)} \exp\left(-\sum_{c \in \mathcal{C}} f_{T(c)}^\theta(\mathbf{x}_c; I) - \sum_{v \in \mathcal{V}} g^\theta(\mathbf{x}_v; I)\right) \quad (2)$$

where X is the labeling of the random field, and \mathbf{x}_c and \mathbf{x}_v are the configurations of clique c and node v respectively. The configuration of a clique or node includes its labeling, and may also include input-dependent latent variables such as image gradient. This formulation is similar to a CRF except that parametric functions over the clique

¹ The time for computing the gradient is linear in clique size for using linear filters, and quadratic in the general case. At the same time, larger cliques also tend to require more iterations.

and node configuration space \mathcal{X}_f and \mathcal{X}_g are used instead of features. The random field becomes Markovian when f is independent of the observed data, i.e. $f_{T(c)}^\theta(\mathbf{x}_c; I) = f_{T(c)}^\theta(\mathbf{x}_c)$ and g is a function only of the observation at a single node, i.e. $g^\theta(\mathbf{x}_v; I) = g^\theta(\mathbf{x}_v; I(v))$.

3 Parameter Estimation

To learn the parameters θ , it is desirable to find the maximum *a posteriori* (MAP) estimate. By applying Bayes' rule and assuming a uniform prior over the parameter space, this is equivalent to finding the maximum likelihood (ML) estimate. Computing the maximum likelihood estimate is nevertheless hard on loopy graphs due to the intractability of the partition function $Z(\theta)$ in $p_\theta(X|I)$. This makes it impossible to use the standard CRF learning scheme, since it is designed for tree-structured graphs where the partition function can be computed efficiently using dynamical programming [12]. Various approaches have been proposed to address this difficulty. Gradient descent methods [18] have been used to obtain a local minimum in the negative log-likelihood space. The expectation over the model is nonetheless intractable to compute and often has to be estimated by MCMC sampling [1,18], by loopy belief propagation [7,19], or approximated using the mode (i.e. MAP labeling) [20]. The last case resembles the perceptron algorithms [21], except that the inference is not exact. As recently proposed in [16], a basis rotation algorithm based on expectation maximization (EM) can be used to learn parameters for filter based image models. This comes from a key observation that the partition function can be kept constant by constraining the parameter vectors to have unit norm. An alternative to maximum likelihood is using discriminative training to optimize for some loss function, typically evaluated on the mode. Such a loss can be minimized by descending along its derivative in the parameter space, when the mode has a closed-form solution [14] (or approximate solution [6]).

Since some approximation must be used, we take the approach of optimizing for the marginal likelihood of the random field cliques, which effectively approximates the global partition function using the product of local partition functions over the cliques. This can be considered as form of piecewise local training [22,23], which minimizes a family of upper bounds on the log partition function. It can be shown that maximizing the marginal likelihood is equivalent to minimizing the Kullback-Leibler (KL) divergence $D_{\text{KL}}(p_0||p_\theta)$ between the empirical distribution p_0 and the model distribution p_θ for each type of cliques. The minimization can be performed using gradient descent with the standard update rule (as in [1])

$$\delta\theta = \eta \left[\left\langle \frac{\partial f_\theta}{\partial \theta} \right\rangle_{p_\theta} - \left\langle \frac{\partial f_\theta}{\partial \theta} \right\rangle_{p_0} \right],$$

where $\langle \cdot \rangle_{p_\theta}$ and $\langle \cdot \rangle_{p_0}$ denote the expectation with respect to the model and the empirical distribution respectively, and η is the learning rate.

Unlike in FoE we do not need to sample, since the model expectation can be computed by summing over all possible values of clique configurations. This computation is inexpensive in our model due to the small clique size. As noted in [1] performance

can be improved by learning an additional weight for the data term, which we also use for our model.

4 Image Denoising

To test the effectiveness of our model, we apply it to the widely studied problem of image denoising. As is conventional in the literature, the image is assumed to be gray-scale and have been corrupted by additive white Gaussian noise of known standard deviation. Since this is a well-defined generative process, we model the data term using the known Gaussian noise model and only the spatial term needs to be estimated. As described above we use 3-cliques since they capture second-order properties. In order to illustrate the importance of these second-order statistics we considered the marginal statistics of the images in the Berkeley dataset [24] that is commonly used in evaluations of such methods. These images show a strong correlation between the distribution of neighboring pairs, suggesting that simple pairwise models are less appropriate (see Figure 2).

We denote clique c of type $C_{s,s,2s}$ as a triplet $(v_{-s}^c, v_0^c, v_{+s}^c)$, where v_0^c is the center node of c , v_{-s}^c is the left node, and v_{+s}^c is the right node. We limit our discussion to horizontal cliques, as the case for vertical ones is essentially the same. Let $d_1(c) = X(v_{+s}^c) - X(v_{-s}^c)$ and $d_2(c) = X(v_{-s}^c) + X(v_{+s}^c) - 2X(v_0^c)$, where X is the labeling of the image. Hence d_1 and d_2 are proportional to the discrete first and second derivatives of the image luminance respectively. In other words, the clique potential couples both first and second order spatial information.

The Lorentzian function has been widely used to model the statistics of natural images (e.g., [25,1,6]). In our case, we use a family of 2-dimensional Lorentzian functions for the spatial term, i.e.

$$f(\mathbf{x}_c) = \alpha \cdot \log(1 + \frac{1}{2}[(\beta_1 d_1)^2 + (\beta_2 d_2)^2]) \tag{3}$$

where $\{\alpha, \beta_1, \beta_2\}$ is the set of parameters for cliques of type $T(c)$. Hence f is intensity-invariant and regulates both the first and the second derivatives of the spatial signal. We

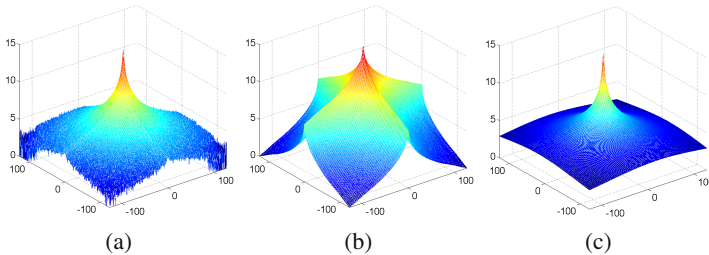


Fig. 2. Frequency (unnormalized, logarithm scale) plotted against gradients of the two neighboring pairs in a linear 3-clique, from the Berkeley dataset [24]. (a) The empirical marginal distribution. (b) The would-be distribution if gradients of the neighboring pairs were independent. (c) The distribution from a fitted Lorentzian cost function.

choose this family since it not only fits the statistics of natural images (Figure 2) but is also able to produce smooth gradient while preserving discontinuities. This form is subtly different from filter based models, such as [1,16], that use a linear combination of functions over filter responses; in our case the first and second order derivatives are coupled, that is both orders of derivatives are inputs to the same non-linear function rather than using a linear combination of separate non-linear functions of each spatial filter.

It has been noted that natural images are self-similar over different spatial scales [26,27]. As a result, cliques with different scales (i.e. edge lengths) all have very similar marginal distributions. This makes the marginals of cliques at different scales highly correlated, which we also observed empirically. Hence using independently collected marginals as the clique potentials is not a good model when dealing with natural scenes. To account for this factor, we reweigh the distribution of smaller-scale cliques according to the marginals of larger-scale ones, so as to make the former learn different trends from what have already been captured by the latter.

4.1 Inference

For denoising, inference can be performed using either belief propagation (BP) [5] or gradient based methods such as limited memory BFGS (L-BFGS) [28]. We experimented with both and found that L-BFGS produces the same quality of results as BP while requiring less running time. Hence the results we report in this paper are based on using L-BFGS. It should be noted, however, that some problems in vision are of a discrete nature and cannot be solved using gradient-based methods. In those cases, discrete optimization techniques such as BP and graph cuts are required.

5 Experimental Results

To evaluate the model for image denoising we used the Berkeley Segmentation Dataset and Benchmark [24] in order to compare the results with previous methods. The models were trained on the training images in the dataset and performance was measured on a subset of the testing images, which are the same as the ones used in [1,6,13,14]. In all the experiments we ran L-BFGS for 20 iterations, which we found to be sufficient for our model.² This is in contrast to large-clique methods, which usually require many hundred iterations to produce results of good quality [1,14].

Table 1 shows the denoising performance of our model along with the results from the FoE model in [1], the steerable random field (SRF) in [13], the Gaussian CRF in [14], and the variational MRF in [6]. This table reports the peak signal-to-noise ratio (PSNR) of each method averaged over the 68 test images (higher is better). These results demonstrate that the performance of our approach is comparable to that of recent top performing random field methods using the standard measure of PSNR. However, as is widely recognized, PSNR does not tell the entire story, thus we also consider some

² We also experimented with conjugate gradient as the optimization method, which achieved the same performance but needs a few more iterations (about 30 as opposed to 20 for L-BFGS).

Table 1. Denoising performance of SLRF measured in peak signal-to-noise ratio (PSNR), higher is better. Results from other random field based denoising methods are shown for comparison. (Bold indicates the best performance among the 3-clique MRF models, asterisk denotes the best overall result, and “–” indicates no published data available.)

Model \ Noise σ	5	10	15	20	25
SLRF, $K=4$	36.90	32.71*	30.39	28.86*	27.73
Local MRF, $K=2$	36.51	32.04	29.81	27.89	26.41
FoE [1]	–	32.67	30.47*	28.79	27.59
GCRF [14]	–	–	–	–	28.04*
Var. MRF [6]	–	–	30.25	–	–
SRF [13]	–	–	–	28.32	–

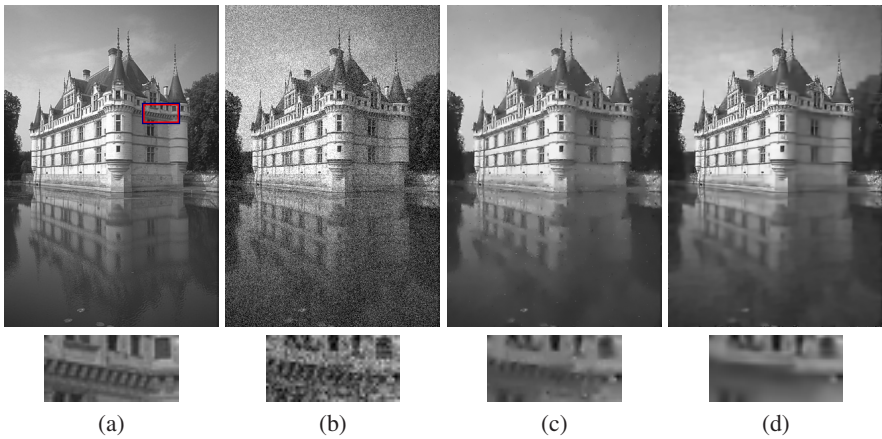


Fig. 3. Denoising output for a medium-texture scene. (a) Original image. (b) Corrupted by Gaussian noise, $\sigma = 25$. (c) Restored using our SLRF model, PSNR = 28.63. (d) Restored using FoE [1], PSNR = 28.72. The magnified view shows that our model, while having comparable PSNR, does a significantly better job at preserving the small and low-contrast structures of the stonework below the windows.

example images in more detail both to show the overall quality and to highlight the extent to which our method removes noise without smearing out the details.

Figures 3 and 4 display sample outputs from our model (in c) and from FoE (in d), illustrating the comparable quality of our method and FoE. In particular our method is able to reproduce image texture without yielding to the visually unpleasant blockiness that other methods using small cliques tend to produce [5,29]. The enlarged regions in each of the images illustrate that our method is able to reproduce fine-scale texture better than the FoE. For instance in the castle image (Fig. 3), the stonework detail below the windows is smoothed out in the FoE but preserved in our model. The textured surface of the rocks in the sheep image (Fig. 4) similarly illustrates the ability of our method to preserve realistic texture while removing noise, rather than over-smoothing. Moreover, our method produces a consistent level of sharpness across the whole image, and, unlike

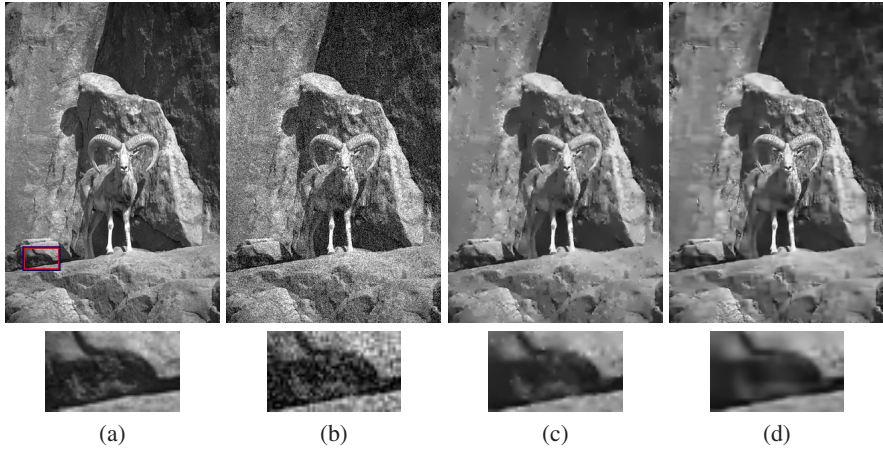


Fig. 4. Denoising output for a high-texture scene. (a) Original image. (b) Corrupted by Gaussian noise, $\sigma = 25$. (c) Restored using our SLRF model, PSNR = 26.02. (d) Restored using FoE, PSNR = 25.55. Again, the detail illustrates that our model not only achieves good PSNR but also produces less over-smoothing.

FoE, does not tend to make high-contrast regions very sharp while low-contrast regions very smooth (Fig. 3 and 4, compare (c) and (d)). This gives the output of our model a more natural look.

Table 1 also shows that the model with long-range edges ($K = 4$) performed better than the local model ($K = 2$), in terms PSNR, and that the difference is most pronounced at high noise levels (e.g. $\sigma = 25$) as would be expected. Even at low noise levels (e.g. $\sigma = 5$), where one would not necessarily expect much help from longer-range connections, the long-range model still slightly outperformed the local model. This suggests that long-range interactions increase robustness of the model without sacrificing fine-scale precision. Figure 5 shows in side-by-side comparison some sample output of the long-range model and the local model. The difference in visual quality between the two emphasizes that longer-range connections are useful and that our simple second-order models are capturing important properties of images, though these are not completely reflected in the PSNR numbers.

In addition to the experiments with artificial Gaussian noise, we also test our model on real-world noisy images. For color images, we simply transform them into YCbCr space and apply the model on each channel separately. In all our experiments, Gaussian white noise is assumed. Although this is suboptimal, we obtain qualitatively good results as can be seen in Figure 6.

These results illustrate that our model utilizing sparse long-range connections achieves state-of-the-art performance when compared with other random field methods for image denoising. Arguably the better preservation of texture and more natural look compared with FoE, without the blocky effects of other local methods, improves upon previous results. Due to the small clique size and hence low complexity, our model is less prone to artifacts, such as the ringing pattern, which occurs more often with

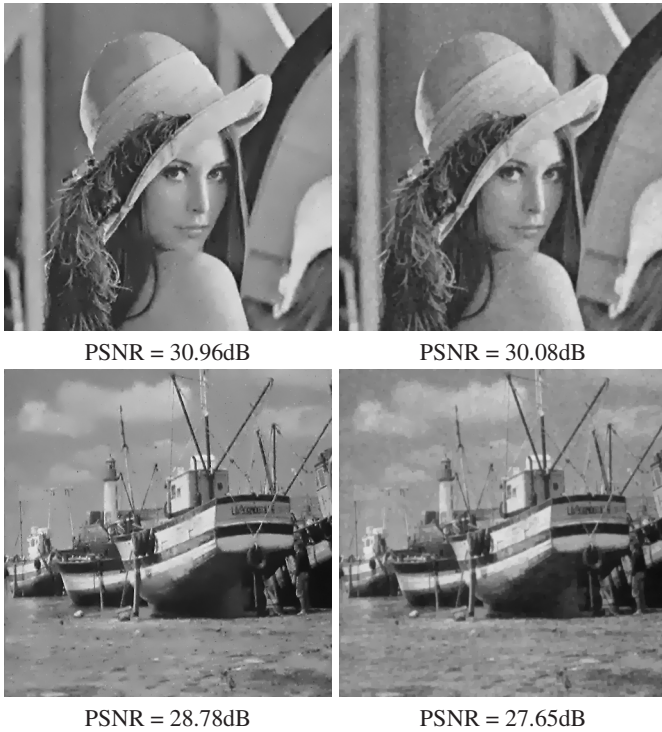


Fig. 5. Comparison of denoising outputs of the long-range and the local models. Input images have Gaussian white noise with $\sigma = 25$ (PSNR = 20.17). Left: Results of the long-range ($K = 4$) model. Right: Results of the local ($K = 2$) model. Observe that the outputs of the local model is blocky and appear tainted while those of the long-range model are smooth and clean.

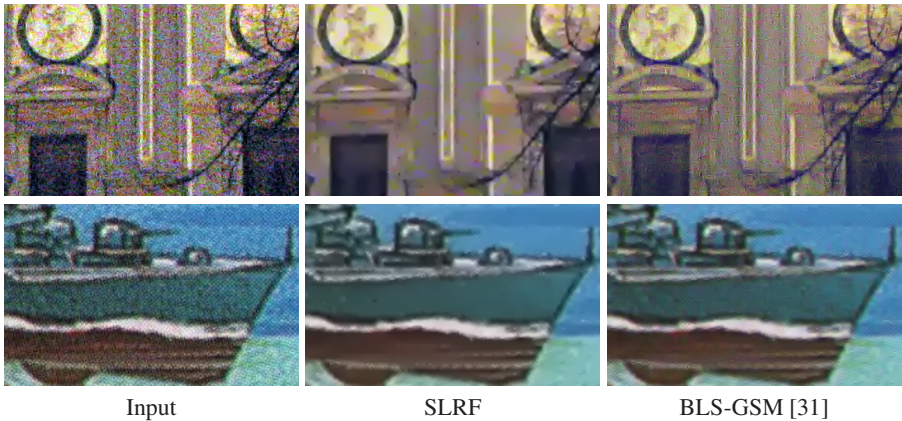


Fig. 6. Results on two real-world noisy images used in [31]. For these two images, our model assumes Gaussian white noise of standard deviation of 50 and 25 respectively. Despite the lack of accurate noise model, the visual quality of the output of our method is at least as good as that of [31].

Table 2. Running time of various image denoising methods

Method	Image size	Processor	Running time (sec.)
SLRF	481×321	Xeon-3.0GHz	3.2
FoE [1]	481×321	Xeon-3.2GHz	376.9
GCRF [14]	481×321	Xeon-3.2GHz	97.8
GSM [30]	256×256	PentiumIII-1.7GHz	approx. 40

higher-order models. The highest PSNR has been achieved by wavelet based methods (e.g. [30,31]); nevertheless, such models tend to produce a larger amount of ringing artifacts.

Finally we compare in Table 2 the running time of our model with those reported for some other methods, including both random field [1,14] and wavelet-based [30]. These results show that our method is a factor of 30 or more faster than the other random field methods and about 10 times faster than the wavelet-based one (note that while the running time in this last case is for a slower processor, the image is also considerably smaller). The speed of our model makes it a practical denoising method even for high-resolution images.

6 Conclusion

We have presented a model which explicitly represents long-range interactions but only uses low-order cliques, thereby enabling much faster optimization than other approaches that rely on high-order cliques. For image denoising this model achieves state-of-the-art PSNR results among random field methods, is better at preserving fine-scale detail, and runs at least an order of magnitude faster. The low complexity nature of the model not only reduces artifacts such as ringing, but also makes it readily interpretable and easy to understand. The small clique size enables the use of efficient approximate global inference algorithms for arbitrary clique potentials, whilst the explicit long-range interactions effectively counters noise-induced local bias. The combination of speed and expressiveness makes it an efficient and robust approach for low-level vision problems in noisy domains.

Acknowledgments

This work was supported in part by NSF grant IIS-0713185.

References

1. Roth, S., Black, M.J.: Fields of experts: A Framework for Learning Image Priors. In: CVPR (2005)
2. Besag, J.E.: Spatial interaction and the statistical analysis of lattice systems. J. Royal Stat. Soc. B 36(2) (1974)

3. Geman, S., Geman, D.: Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *PAMI* 6(6), 721–741 (1984)
4. Szeliski, R.: Bayesian modeling of uncertainty in low-level vision. *IJCV* 5(3) (1990)
5. Lan, X., Roth, S., Huttenlocher, D.P., Black, M.J.: Efficient belief propagation with learned higher-order Markov random fields. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) *ECCV 2006*. LNCS, vol. 3951. Springer, Heidelberg (2006)
6. Tappen, M.F.: Utilizing variational optimization to learn Markov random fields. In: *CVPR* (2007)
7. Murphy, K., Weiss, Y., Jordan, M.: Loopy belief propagation for approximate inference: An empirical study. In: *UAI* (1999)
8. Kolmogorov, V.: Convergent tree-reweighted message passing for energy minimization. *PAMI* 28(10) (2006)
9. Wainwright, M.J., Jaakkola, T.S., Willsky, A.S.: Map estimation via agreement on (hyper)trees: Message-passing and linear programming approaches. Technical Report UCB/CSD-03-1269, EECS, UC Berkeley (2003)
10. Boykov, Y., Veksler, O., Zabih, R.: Fast approximate energy minimization via graph cuts. In: *ICCV* (1999)
11. Szeliski, R., Zabih, R., Scharstein, D., Veksler, O., Kolmogorov, V., Agarwala, A., Tappen, M.F., Rother, C.: A comparative study of energy minimization methods for Markov random fields. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) *ECCV 2006*. LNCS, vol. 3951. Springer, Heidelberg (2006)
12. Lafferty, J., McCallum, A., Pereira, F.: Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In: *ICML* (2001)
13. Roth, S., Black, M.J.: Steerable random fields. In: *ICCV* (2007)
14. Tappen, M.F., Liu, C., Adelson, E.H., Freeman, W.T.: Learning gaussian conditional random fields for low-level vision. In: *CVPR* (2007)
15. Geman, S., Graffigne, C.: Markov random field image models and their applications to computer vision. In: *Intl. Congress of Mathematicians* (1986)
16. Weiss, Y., Freeman, W.F.: What makes a good model of natural images? In: *CVPR* (2007)
17. Gimel'farb, G.L.: Texture modeling by multiple pairwise pixel interactions. *PAMI* 18(11) (1996)
18. Hinton, G.E.: Training products of experts by minimizing contrastive divergence. *Neur. Comput.* 14(8) (2002)
19. Yedidia, J.S., Freeman, W.T., Weiss, Y.: Generalized belief propagation. In: *NIPS* (2000)
20. Scharstein, D., Pal, C.: Learning conditional random fields for stereo. In: *CVPR* (2007)
21. Collins, M.: Discriminative training methods for hidden Markov models: Theory and experiments with perceptron algorithms. In: *EMNLP* (2002)
22. Sutton, C., McCallum, A.: Piecewise training of undirected models. In: *UAI* (2005)
23. Sutton, C., Minka, T.: Local training and belief propagation. Technical Report TR-2006-121, Microsoft Research (2006)
24. Martin, D., Fowlkes, C., Tal, D., Malik, J.: A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: *ICCV* (2001)
25. Huang, J., Mumford, D.: Statistics of natural images and models. In: *CVPR* (1999)
26. Freeman, W.T., Pasztor, E.C., Carmichael, O.T.: Learning low-level vision. *IJCV* 40(1) (2000)
27. Srivastava, A., Lee, A., Simoncelli, E., Zhu, S.: On advances in statistical modeling of natural images. *Journal of Mathematical Imaging and Vision* (2003)

28. Nocedal, J.: Updating quasi-newton matrices with limited storage. *Mathematics of Computation* 35, 773–782 (1980)
29. Felzenszwalb, P.F., Huttenlocher, D.P.: Efficient belief propagation for early vision. *IJCV* 70(1) (2006)
30. Portilla, J.: Blind non-white noise removal in images using gaussian scale mixtures in the wavelet domain. In: *Benelux Signal Processing Symposium* (2004)
31. Portilla, J., Strela, V., Wainwright, M.J., Simoncelli, E.P.: Image denoising using scale mixtures of gaussians in the wavelet domain. *IEEE Trans. Imag. Proc.* 12(11) (2003)