An Implicit Markov Random Field Model for the Multi-scale Oriented Representations of Natural Images

Siwei Lyu
Computer Science Department
SUNY Albany, Albany, NY 12222, USA
lsw@cs.albany.edu

Abstract

In this paper, we describe a new Markov random field (MRF) model for natural images in multi-scale oriented representations. The MRF in this model is specified with the singleton conditional densities (the density of one subband coefficient given its Markovian neighbors), while the clique potentials and joint density of this model are implicitly *defined. The singleton conditional densities are chosen to have maximum entropy and consistent with observed statistical properties of natural images. We then describe parameter learning for this model, and a sparse prior to choose optimal model structure. Using this model as image prior, we develop an iterative image denoising method, and a solution to restoring images with missing blocks of subband coefficients.*

1. Introduction

Statistical models for natural images are important tools in image processing and computer vision, which play essential roles in problems such as denoising, inpainting, super-resolution and texture synthesis. Desirable properties of such models include:

- **consistency** with observed statistical properties of natural images,
- **scalability** in modeling images without constraint on image sizes, and
- **efficiency** of computation in parameter learning, sampling and inference.

A case in point are the Gaussian models, which can capture second order statistics of natural images, and afford efficient computation in high dimensional data space. However, studies in the recent two decades have showed that Gaussian models are not good models for natural images. Especially, when transformed into multi-scale oriented representations (e.g., wavelets), natural images exhibit striking non-Gaussian marginal statistics and statistical dependencies beyond second order. For this reason, there have been several non-Gaussian models that can match statistical properties of natural images over small blocks of pixels or subband coefficients[22, 11, 9, 25]. These models are easy to compute and understand, and they have achieved state-of-the-art performance in problems such as image denoising [17]. The downside, however, is that there is no straightforward extension of such models to images with sizes beyond small blocks. In parallel, there have also been significant advances in using global probabilistic models such as Markov random fields (MRFs) for natural image [10, 12, 28, 8, 21]. In an MRF, local statistical behaviors are embedded in the global image domain in a consistent manner, and can propagate properly across different image locations. Their main disadvantages are that computation in MRF models is usually much more expensive than that of the patch based models, and the construction of MRF models usually compromise on matching statistical properties of natural images for computation efficiency.

In this paper, we introduce a new MRF model for natural images in the multi-scale oriented image representations. Motivated by higher-order statistical regularities of natural images, this model is constructed with specifications of all singleton conditional densities (the density of one subband coefficient given its Markovian neighbors), which can uniquely determine the joint density of an MRF. The singleton conditional densities are chosen to have maximum entropy and consistent with observed statistical properties of natural images. Because this MRF model is not directly defined with clique potentials, or equivalently, the joint probability density function, we call it an *implicit* MRF model. We show that the parameters of the implicit MRF model can be estimated using the maximum pseudo-likelihood principle. Furthermore, the structure of the MRF can be determined with a sparse inducing prior over model parameters. To evaluate its effectiveness in modeling natural image statistics, we apply the implicit MRF model to the Bayesian solution of image denoising and the restoration of images with damaged blocks of subband coefficients.

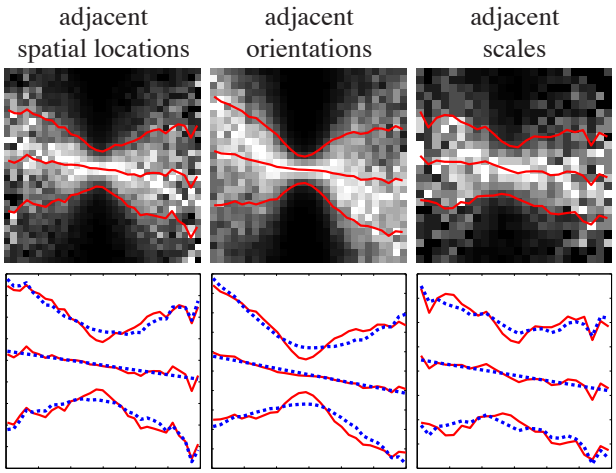


Fig. 1. Top: Conditional histogram of adjacent pairs of subband coefficients of a natural image in a multi-scale oriented representation, with different spatial locations, orientations or scales. Overlaid on the conditional densities, as well as in the corresponding plots in the bottom row as red solid curves, are the estimated means (in the middle) and the range of one standard deviation. **Bottom**: Linear model fittings of the conditional means and standard deviations as blue dashed curves.

2. Statistical Properties of Natural Images in Multi-scale Oriented Representations

Over the past two decades, multi-scale oriented representations have been widely adopted in image processing and computer vision, notable examples include wavelets [15], ICA bases [1], curvelets [24] and steerable pyramids [23]. Under these representations, images are projected onto basis functions that are localized in space, frequency and local orientations. One major advantage of the multi-scale oriented representations over the pixel or frequency representations is that they are very effective to reveal non-Gaussian statistical properties of natural images. For instance, the coefficients within one subband (i.e., the projections of an image on the basis functions within the same range of spatial frequencies and local orientations) tend to have heavy-tail non-Gaussian marginal distributions [6]. More importantly, there are strong higher-order statistical dependencies across space, orientations and scales in the multi-scale oriented decompositions of natural images [26, 5].

These higher-order statistical dependencies can be more clearly illustrated if we examine the conditional distributions of two coefficients, (x_1, x_2), that are adjacent in spatial locations, orientations or scales, as shown in the top row of Fig.1[1]. Specifically, the first column is for the coefficient pairs from the same subband with spatial separation of two samples. The second column corresponds to two coefficients of the same spatial location and scale but with a difference in orientation of $\pi/8$. The third column corresponds to two coefficients of the same spatial location and orientation but from two adjacent scales. Each vertical slice of the images in the top row represents the conditional density $p(x_1|x_2)$ for a particular value of x_2, and the brightness in the images are proportional to probability. For better visualization effects, each vertical slice is normalized independently to fill the full range of intensities. Overlaid on the conditional densities, as well as in the corresponding plots in the bottom row of Fig.1 as red solid curves, are the estimated means (in the middle) and the ranges of one standard deviation for $p(x_1|x_2)$. These conditional histograms have shapes similar to a "bow-tie" [5], implying that x_1 and x_2 are dependent - otherwise the conditional densities should be the same across different x_2. Furthermore, such dependencies are non-Gaussian, as both the conditional means and the conditional standard deviations vary, in contrast to the Gaussian case where the conditional standard deviations should be a constant.

The dependencies between x_1 and x_2 can be quantified by fitting linear models to the mean and variance of $p(x_1|x_2)$ using values of x_2 and x_2^2, as $\mathcal{E}(x_1|x_2) \approx ax_2$, and $\text{var}(x_1|x_2) \approx b + cx_2^2$, where a, b and c can be estimated from data. As shown in the bottom row of Fig.1 with blue dashed curves, these linear models provide very good fittings to the observed conditional means and variances. Note that the dependencies in the conditional means may be removed by a linear whitening transform or with a de-correlated representation (e.g., the orthogonal wavelets), where the conditional means are reduced to zero. On the other hand, no simple linear transform can remove the higher-order non-Gaussian dependencies as presented in the conditional standard deviations [14].

3. Implicit MRF Model for Natural Images in Multi-scale Oriented Representations

The higher-order statistical dependencies as shown in Fig.1 also hold for blocks of subband coefficients at adjacent locations, orientations and scales, where the mean of one coefficient given a set of neighboring coefficients is approximately a linear combination of the neighboring coefficients, and its conditional variance is a linear combination of the squared neighboring coefficients plus a constant [5]. We would like to construct a Markov random field (MRF) model over all subband coefficients that can capture these properties of natural images, where one coefficient is independent of the rest of the decomposition given its neighbors in a generalized neighborhood of adjacent coefficients in space, orientation and scale (known as the Markov blanket, and denoted as $\mathcal{N}(i)$) as: $p(x_i|x_{j,j\neq i}) = p(x_i|x_{j,j\in\mathcal{N}(i)})$. These Markov blankets correspond to cliques in the equiv-

[1]The results are obtained using the steerable pyramid [23] as image representation. Similar observations have also been reported under other decompositions [11, 9, 5].

alent graph structure of the MRF. Most MRF-based image statistical models [10, 12, 28, 8, 21] were constructed with specifications of clique potentials, as they are equivalent to the joint density of the MRF according to the Hamersley-Clifford theorem [27]. Here, on the other hand, the statistical properties of natural images in the multi-scale oriented representations are given in terms of conditional statistics, it may be easier to find singleton conditional densities, $p(x_i|x_{j,j\in\mathcal{N}(i)})$, that satisfy these constraints. According to the Brook's Lemma [27], if we can specify all such singleton conditional densities for each coefficient x_i, they can uniquely determine the joint density of the MRF.

For the sake of subsequent description, let us denote $\mu_i = \mathcal{E}(x_i|x_{j,j\in\mathcal{N}(i)})$ and $\sigma_i^2 = \mathrm{var}(x_i|x_{j,j\in\mathcal{N}(i)})$ as the conditional mean and variance of a coefficient x_i given its generalized neighborhood, respectively. The observations on natural images suggest that μ_i and σ_i^2 have a linear relation with the value of the neighbors, as:

$$\mu_i = \sum_{j\in N(i)} a_j x_j \ \text{ and } \ \sigma_i^2 = b + \sum_{j\in N(i)} c_j x_j^2. \qquad (1)$$

Note that Eq.(1) only specifies the mean and variance of conditional density $p(x_i|x_{j,j\in N(i)})$, and there are many density functions that can have the same first and second order-statistics. To obtain a unique solution, we employ the maximum entropy principle [7] . which advocates density $p(x_i|x_{j,j\in N(i)})$ that has the maximum differential entropy while satisfying Eq.(1). The maximum entropic density has the intuitive interpretation that besides Eq.(1) it has no other hidden assumptions about data. With constraints on the mean and the variance, the maximum entropic density is a Gaussian with the given mean and variance [7], which, in this case, yields:

$$p(x_i|x_{j,j\in\mathcal{N}(i)}) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{(x_i-\mu_i)^2}{2\sigma_i^2}\right). \qquad (2)$$

A complete set of singleton conditional densities can be specified by repeating this for every coefficient x_i, which in turn determine a joint MRF model. Note that though this MRF has Gaussian singleton conditional densities, it is not jointly Gaussian, due to the dependencies of the conditional variances on the neighboring coefficient magnitudes.

3.1. Learning

To use the implicit MRF model, we need to determine the model structure (i.e., the formation of the Markov blanket) and the model parameters. To reduce the number of parameters, we assume that the implicit MRF model has homogeneous model structure and parameters, i.e., a fixed set of parameters, $\{a_i, b, c_i\}$, and the same Markov blanket structure, independent of the spatial location.

Given the model structure, the model parameters, $\{a_i, b, c_i\}$, are typically learned with maximum likelihood. Unfortunately, two reasons make direct maximum likelihood learning for the implicit MRF model very challenging. First, it requires the computation of the normalizing partition function in the joint density, which, for high dimensional data, is usually very hard. More important, the proposed implicit MRF model is defined with the ensemble of all singleton conditional densities. To form the joint density required in maximum likelihood learning, we may have to integrate in high-dimensional space, which makes the optimization even harder. For these reasons, in this work, we adopt *maximum pseudo-likelihood* (MPL) learning [2], which maximizes the product of all singleton conditional densities as: $\max_\theta \prod_i p(x_i|x_{j,j\neq i};\theta)$, or equivalently $\sum_i \log p(x_i|x_{j,j\in N(i)};\theta)$, with regards to θ, where $\theta = (\vec{a}, b, \vec{c})$ combines all model parameters, and $\vec{a}$ and $\vec{c}$ contain the linear weights in Eq.(1) of a given order, respectively. It has been shown that under some mild conditions MPL can achieve unbiased estimation of the true model parameters [27]. For the implicit MRF model, using Eq. (2), the MPL objective function is:

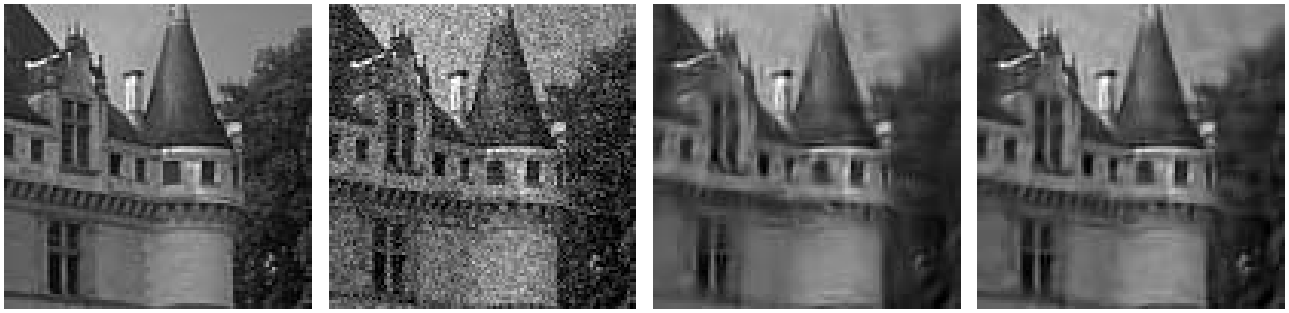
$$\sum_i \left[\log\sigma_i^2 + \frac{(x_i-\mu_i)^2}{\sigma_i^2}\right]$$
$$= \sum_i \left[\log(b + \sum_{j\in N(i)} c_j x_j^2) + \frac{(x_i - \sum_{j\in N(i)} a_j x_j)^2}{b + \sum_{j\in N(i)} c_j x_j^2}\right] \qquad (3)$$

with constraint $b + \sum_{j\in N(i)} c_j x_j^2 > 0$. It can be shown that the objective function given in Eq.(3) and its constraints are quasi-concave, which can be solved efficiently with convex programming with a guarantee for the global optimum [20].

The model structure is determined by augmenting the parameter learning procedure with a penalty on neighboring coefficients whose contribution to the overall objective function, Eq.(3), is small. Specifically, we start with an over-estimation of the neighborhood $\tilde{N}(i)$, then an objective function formed by Eq.(3) with a penalty term on the "sparseness" of the model parameters, as:

$$\underset{\vec{a},b,\vec{c}}{\mathrm{argmin}} \sum_i \left[\log\sigma_i^2 + \frac{(x_i-\mu_i)^2}{\sigma_i^2}\right] + \eta_a|\vec{a}| + \eta_c|\vec{c}| \qquad (4)$$

with the same constraint. $|\cdot|$ stands for l_1 vector norm, and η_a and η_c are meta-parameters controlling the penalties on non-sparseness of parameter $\vec{a}$ and $\vec{c}$, respectively. This augmented objective function is still quasi-concave and thus can be solved with convex programming methods. With the optimal solution, we discard values in vector $\vec{a}$ and $\vec{c}$ that are smaller than pre-given thresholds ϵ_a and ϵ_c, and the neighborhood is formed as $N(i) = \{j|(j\in\tilde{N}(i)) \wedge (\vec{a}_j > \epsilon_a) \wedge (\vec{c}_j > \epsilon_c)\}$. With $N(i)$ fixed, the model parameters are estimated by optimizing Eq.(3).



clean noisy ($\sigma = 20$) GSM[17] (28.37dB) ICM-MAP (28.68dB)

Fig. 2. Image denoising results using GSM model based method [17] and the ICM-MAP method described in this paper for noise level $\sigma = 20$. See texts for more detail.

4. ICM-MAP Image Denoising

In this section we describe an iterative Bayesian image denoising algorithm based on the learned implicit MRF model for natural images in multi-scale oriented representations. The observed image is assumed to be corrupted with *i.i.d* Gaussian noise of zero mean and known variance . We first transform the noise-corrupted image into a multi-scale oriented domain, where the optimal estimation of the "clean" image is estimated in a Bayesian inference framework, using the implicit MRF model learned from a set of natural images as prior. Denote $\vec{x}$, $\vec{w}$ and $\vec{y}$ as the "clean" image, the additive Gaussian noise, and the noisy image in the multi-scale oriented representation, respectively. As the transform from pixels to the multi-scale oriented representation is linear, we have $\vec{y} = \vec{x} + \vec{w}$. However, in a correlated or over-complete representation, white Gaussian noise in the pixel domain will become correlated. Therefore $\vec{w}$, or the conditional density of $\vec{y}$ given $\vec{x}$ is given as,

$$p(\vec{y}|\vec{x}) = \frac{1}{\sqrt{2\pi|C_w|}} \exp\left(-\frac{1}{2}(\vec{y}-\vec{x})^T C_w^{-1}(\vec{y}-\vec{x})\right), \tag{5}$$

with C_w being the covariance matrix that obtained from the linear transform of the multi-scale oriented representation. The denoising procedure is performed as a Bayesian inference of $\vec{x}$ from $\vec{y}$. The restored image is obtained by inverting the linear transform from the multi-scale oriented domain back to the pixel domain. The performance of denoising results is evaluated with the peak-signal-to-noise-ratio (PSNR), defined as $20\log_{10}(255/\sigma_e)$, with σ_e the standard deviation (computed by averaging over spatial position) of the difference between the restored image and the original image.

To use the implicit MRF model for denoising, the inference of $\vec{x}$ given $\vec{y}$ is formulated as finding the *maximum a posterior* (MAP) solution, as $\operatorname{argmax}_{\vec{x}} p(\vec{x}|\vec{y}) = \operatorname{argmax}_{\vec{x}} p(\vec{y}|\vec{x})p(\vec{x}) = \operatorname{argmax}_{\vec{x}} \log p(\vec{y}|\vec{x}) + \log p(\vec{x})$. Directly optimizing the MAP objective function using gradient based methods may be very difficult, due to the high dimensionality and the need of a joint density for $\vec{x}$. Instead, taking advantage of the singleton conditional densities in the implicit MRF model, we use *iterative conditional mode* (ICM) [3], which iteratively optimizes each x_i in the MAP objective function, while fixing the current values of the remaining $x_{j,j\neq i}$, as:

- Start with an initial value for $\vec{x}^{(0)}$, and set $t = 1$.
- Repeat until convergence
  - Repeat for all i
    - Compute the current estimation for x_i, as $x_i^{(t)} = \operatorname{argmax}_{x_i} \log p(x_1^{(t)}, \cdots, x_{i-1}^{(t)}, x_i, x_{i+1}^{(t-1)}, \cdots, x_d^{(t-1)}|\vec{y})$.
  - $t \leftarrow t + 1$.

The ICM algorithm is guaranteed to converge to a *local* maximum of the MAP objective function [3]. Each step updating x_i reduces to a one-dimensional optimization problem that can be solved easily. Specifically, we have

$$\begin{aligned}
&\operatorname*{argmax}_{x_i} \log p(\vec{y}|\vec{x}) + \log p(\vec{x}) \\
=\ &\operatorname*{argmax}_{x_i} \log p(\vec{y}|\vec{x}) + \log p(x_1, \cdots, x_{i-1}, x_i, x_{i+1}, \cdots, x_n) \\
=\ &\operatorname*{argmax}_{x_i} \log p(\vec{y}|\vec{x}) + \log p(x_i|x_{j,j\in N(i)}) + \log p(x_{j,j\in N(i)}).
\end{aligned}$$

We can drop $\log p(x_{j,j\in N(i)})$ in the last step of the above equation as it is a constant with regards to x_i. Also, note that we use the Markovian property to substitute $\log p(x_i|x_1, \cdots, x_{i-1}, x_{i+1}, \cdots, x_n)$ with $\log p(x_i|x_{j,j\in N(i)})$. With Eq.(5) and Eq.(2), dropping constant terms with regards to x_i, and denoting $w_{ij} = (C_w^{-1})_{ij}$, and $\sigma_w^2 = \frac{1}{w_{ii}}$, optimizing $\log p(\vec{y}|\vec{x}) + \log p(x_i|x_{j,j\in N(i)})$ with regards to x_i is equivalent to finding

$$\operatorname*{argmin}_{x_i} \sum_{i\neq j} w_{ij}(x_i - y_i)(x_j - y_j) + \frac{(x_i - y_i)^2}{2\sigma_w^2} + \frac{(x_i - \mu_i)^2}{2\sigma_i^2}.$$

Now taking the derivative of the above function with regards to x_i and setting the result to zero yield

$$x_i = \frac{\sigma_w^2\sigma_i^2}{\sigma_w^2 + \sigma_i^2}\left(\frac{y_i}{\sigma_w^2} + \frac{\mu_i}{\sigma_i^2} - \sum_{i\neq j} w_{ij}(x_j - y_j)\right). \tag{6}$$

A special case of this algorithm, when a de-correlated multi-scale oriented representation such as orthogonal wavelet is used where $\mu_i = 0$ and $w_{ij} = 0$, Eq.(6) reduces to a classical Wiener filtering in removing Gaussian noise from a Gaussian source, as: $x_i = \frac{\sigma_i^2}{\sigma_w^2+\sigma_i^2} y_i$. However, subsequent changes to the values of other coefficients update the estimation of conditional variances of x_i. Thus the iterative ICM-MAP denoising algorithm can be interpreted as alternating between estimation of local signal variance σ_i^2 using Eq. (1) and removing noise based on the estimated variance and a local conditional Gaussian image model, Eq.(2). A similar two-step denoising method was described in [13], though there the algorithm was described in the pixel domain, with two important differences: (1) there is no explicit statistical model to relate with image statistics and (2) the algorithm runs with only one iteration.

4.1. Experiment

We evaluate the performance of the ICM-MAP denoising method based on the implicit MRF model with 20 grayscale images converted from the Berkeley image database [16]. Noisy images were generated by adding *i.i.d*. Gaussian noise with zero mean and standard deviation of $\sigma = 10, 20, 50$. As it has been shown that over-complete representation is particular effective for removing noise [19], we chose the steerable pyramid [23] with 8 orientations and 4 scales as the multi-scale oriented image representation. The resulting decomposition of an image is approximately 11 times over-complete relative to the original image size.

The implicit MRF model for "clean" images in the over-complete steerable pyramid domain is then learned in an off-line manner with another set of 20 images from the Berkeley image database. Specifically, we ran the learning algorithm as described in Section 3.1 on the steerable pyramid decomposition of the training images to find the optimal Markovian neighborhood and model parameters. The initial Markovian neighborhood for a subband coefficient for model structure learning was chosen as:

- spatial neighbors in the 5×5 surrounding block in the same subband,
- 5×5 blocks of the corresponding spatial locations from the two adjacent orientation subbands,
- 5×5 blocks of the corresponding spatial locations from the subband of one scale higher.

As the statistical properties described in Eq.(1) apply only to the band-pass subband coefficients, in this work, we treat the coefficients in the residual low-pass subband as constants. As the additive Gaussian noise has very low energy in the low-pass frequency range, this should not significantly affect the denoising performance. Fig.3 illustrated the initial neighborhood setting, as well as the result

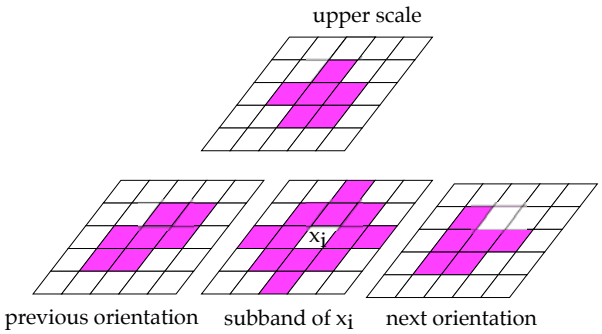


Fig. 3. Initial (blank) and refined (filled) neighborhood structure for the implicit MRF image model for one subband coefficient x_i.

of structure learning (in shade). With the model structure found, the MPL parameter learning was used to find the optimal model parameters. In our experiments, the off-line learning ran on average about two hours to complete, using MATLAB based convex programming package CVX[2].

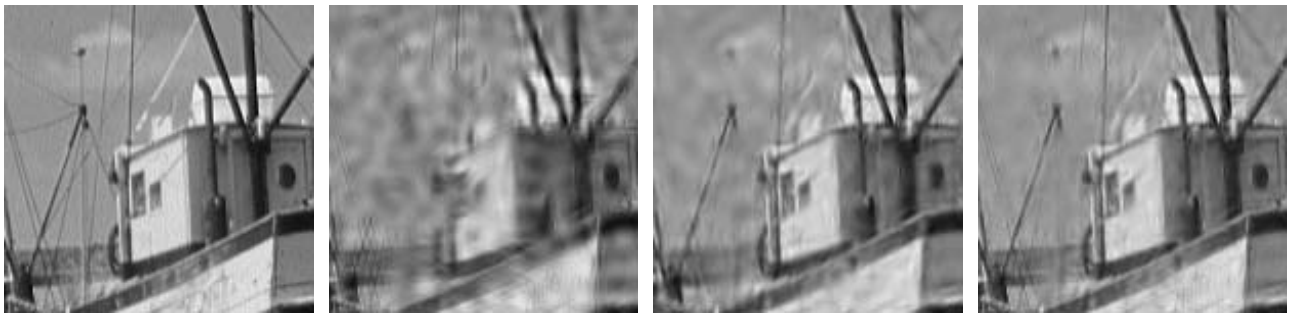
Table 1 shows the denoising results averaged over the 20 testing images, and compared with the `wiener2` function in MATLAB, which implements the algorithm in [13], as well as a state-of-the-art denoising method based on a GSM model for blocks of steerable pyramid coefficients [17]. The results in Table 1 suggest that the proposed ICM-MAP denoising method based on the implicit MRF model achieved better performances measured with PSNR. Fig.2 shows the denoising results of a region of one testing image. As the implicit MRF model is able to capture the local statistical dependencies and propagates such dependencies properly to the extend of whole image, the ICM-MAP denoised image tends to have smoother contours and less noise residuals.

5. Restoration of Corrupted Subband

In this section, the implicit MRF model is applied to restore images corrupted by random missing blocks of subband coefficients. This is a practical problem in image transmission. Encoding image in the multi-scale oriented domain (e.g., wavelet) has many advantages over the DCT based JPEG compression [18]. However, when the compressed image subbands are transmitted, they are susceptible to the transmission errors, which usually cause corruption or missing of small blocks of adjacent coefficients. To recover the original image, we need estimate the missing coefficients based on the intact ones.

We denote the missing coefficient blocks as $\mathcal{C}$, and the intact coefficients adjacent to $\mathcal{C}$ as $\partial\mathcal{C}$. We assume that coefficients in $\mathcal{C}$ are conditionally independent from the rest of the image given the coefficients in $\partial\mathcal{C}$, i.e., $\forall i \in \mathcal{C}$, $N(i) \subset \mathcal{C} \bigcup \partial\mathcal{C}$. As in denoising, we could use the MAP criterion to find the missing coefficients, as: $\text{argmax}_{\{x_i, i\in\mathcal{C}\}} \log p(\{x_i, i \in C\}|\{x_j, j \in \partial\mathcal{C}\})$, and use

[2]CVX is available at http://www.stanford.edu/boyd/cvx/index.html.



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

Fig. 4. Restoration of corrupted subband. (a) Original image. (b) Reconstructed Image with missing blocks of coefficients being replaced with zeros. (c) Reconstructed image using recursive linear regression as in [18]. (d) Reconstructed image using estimations of the missing subband coefficient by optimizing Eq.(7).

methods	$\sigma = 10$	$\sigma = 20$	$\sigma = 50$
MATLAB `wiener2`	29.08	28.44	22.15
GSM [17]	34.21	32.07	27.43
ICM-MAP	**34.86**	**32.48**	**28.03**

Table 1. Average denoising performance measured with PSNR (dB) of different methods on the testing images used in this paper.

methods	5%	10%	20%
zero padding	27.45	25.98	20.23
iterative linear regression	30.11	27.42	22.76
using Eq.(7)	**32.39**	**30.01**	**26.27**

Table 2. Average performance measured with PSNR (dB) of restoration of images with missing subband coefficients of different fractions.

ICM to iteratively find a solution. However, in this case, this process reduces to updating the conditional mean of each singleton conditional density. As they are linear combinations of the neighboring coefficients, Eq.(1), this is equivalent to a recursive linear regression of the missing coefficient using its neighbors [18]. One problem is that no consideration is given to the uncertainties in the conditional variances.

For this reason, we use an alternative objective function, which is used in the MPL learning of model parameters, where $x_i, i \in \mathcal{C}$ are optimized with the model parameter being fixed, as:

$$\underset{x_i, i\in\mathcal{C}}{\operatorname{argmax}} \sum_{i\in C} \left[\log(b + \sum_{j\in N(i)} c_j x_j^2) + \frac{(x_i - \sum_{j\in N(i)} a_j x_j)^2}{b + \sum_{j\in N(i)} c_j x_j^2} \right]. \tag{7}$$

The neighboring coefficients of x_i are split into two non-intersecting groups, one using the set of other missing coefficients, $x_j, j \in N(i) \bigcap \mathcal{C}$, and the other using uncorrupted coefficients $x_j, j \in N(i) \bigcap \partial\mathcal{C}$. The optimal solution can be found similar to the ICM-MAP denoising method by an iterative optimization of x_i with current estimates of other coefficients in $\mathcal{C}$.

To evaluate the performance of the proposed method, we simulate the corruption of subband coefficients due to transmission error as missing of random blocks of coefficients. The multi-scale oriented representation is steerable pyramid, where the implicit MRF model is learned using algorithms in Section 4.1. The fraction of missing coefficients in the overall subband coefficients was controlled to be $10\%, 20\%$ and 40%. We use bordering region $\partial\mathcal{C}$ as 3×3 extensions of the region corresponding to $\mathcal{C}$. Shown in Table 2 are the restoration performance for different percentage of missing coefficients evaluated with PSNR. For comparison, we compared the result of a simple zero-padding (i.e., replaced the missing coefficients with zeros), the iterative linear regression as in [18], and the proposed algorithm. Fig. 4 shows an example of the results of the restoration algorithm. It can be seen that the proposed MPL restoration restores better structures in the subband compared to the recursive linear regression.

6. Discussion and Future Work

In this paper, we described an implicit Markov random field model for natural images in the multi-scale oriented domain. Unlike many MRF-based image statistical models, this MRF model is constructed by the specification of all singleton conditional densities, which are the maximum entropic densities satisfying the empirically observed statistical dependencies of natural images. We give efficient procedures to learn the model parameters and structure, and show simple algorithms of image denoising and restoration of corrupted subband coefficients based on this model.

There are several directions that we would like to further extend this work. First, we are working on obtaining the joint density from the singleton conditionals, with which, it is easier to establish relations of the implicit MRF model to existing models. Second, we have focused on local statistical characteristics of natural images to construct the global MRF model. However, recent works [4] have shown that natural images also have *non-local* statistical dependencies,

i.e., similar structures tend to appear at different and usually not adjacent locations (e.g., textures). We are thus interested in finding models that can account for local as well as non-local image statistics. Finally, we are also working on applications of the proposed implicit MRF image models to other vision and image processing problems, e.g., compression and super-resolution.

References

[1] A. J. Bell and T. J. Sejnowski. The 'independent components' of natural scenes are edge filters. *Vision Research*, 37(23):3327–3338, 1997.

[2] J. Besag. Spatial interaction and the statistical analysis of lattice systems (with discussion). *Journal of the Royal Statistical Society, Series B*, 36(2):192–225, 1974.

[3] J. Besag. On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society, Series B*, 48:259–302, 1986.

[4] A. Buades, B. Coll, and J. Morel. A non local algorithm for image denoising. In *IEEE Int. Conf. on Computer Vision and Pattern Recognition (CVPR 2005) .vol. 2*, San Diego, CA, 2005.

[5] R. W. Buccigrossi and E. P. Simoncelli. Image compression via joint statistical characterization in the wavelet domain. *IEEE Transactions on Image Processing*, 8(12):1688–1701, 1999.

[6] P. Burt and E. Adelson. The Laplacian pyramid as a compact image code. *IEEE Transactions on Communication*, 31(4):532–540, 1981.

[7] T. Cover and J. Thomas. *Elements of Information Theory*. Wiley-Interscience, 2nd edition, 2006.

[8] W. T. Freeman, E. C. Pasztor, and O. T. Carmichael. Learning low-level vision. *International Journal of Computer Vision*, 40(1):25–47, 2000.

[9] P. Gehler and M. Welling. Products of "edge-perts". In Y. Weiss, B. Schölkopf, and J. Platt, editors, *Advances in Neural Information Processing Systems (NIPS)*, pages 419–426. MIT Press, Cambridge, MA, 2006.

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

[11] J. Huang and D. Mumford. Statistics of natural images and models. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, volume 1, pages 541–547, 1999.

[12] F. Jeng and J. Woods. Compound gauss-markov random fields for image estimation. *IEEE Transaction on Signal Processing*, 39(3):683–697, 1991.

[13] J. S. Lee. Digital image enhancement and noise ltering digital image enhancement and noise filtering by use of local statistics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2:165–168, March 1980.

[14] S. Lyu and E. P. Simoncelli. Reducing statistical dependencies in natural signals using radial Gaussianization. In *Adv. Neural Information Processing Systems 21*, volume 21, Cambridge, MA, May 2009. MIT Press.

[15] S. G. Mallat. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11:674–693, July 1989.

[16] 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. In *Proc. 8th Int'l Conf. Computer Vision*, volume 2, pages 416–423, July 2001.

[17] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli. Image denoising using a scale mixture of Gaussians in the wavelet domain. *IEEE Trans Image Processing*, 12(11):1338–1351, November 2003.

[18] F. A. Pujol, H. Mora, J. L. Sánchez, and A. Jimeno. Ezw-based image compression with omission and restoration of wavelet subbands. In *Progress in Pattern Recognition, Image Analysis and Applications*, Lecture Notes in Computer Science, pages 134–141. Springer Berlin / Heidelberg, 2008.

[19] M. Raphan and E. P. Simoncelli. Optimal denoising in redundant bases. In *Proc 14th IEEE Int'l Conf on Image Proc*, San Antonio, TX, September 2007. IEEE Computer Society.

[20] R. T. Rockefeller. *Convex Analysis*. Princeton Univ. Press, Princeton, NJ, 1970.

[21] S. Roth and M. Black. Fields of experts: A framework for learning image priors. In *IEEE Conference on Computer Vision and Patten Recognition (CVPR)*, volume 2, pages 860–867, 2005.

[22] E. P. Simoncelli and E. H. Adelson. Noise removal via Bayesian wavelet coring. In *Proc 3rd IEEE Int'l Conf on Image Proc*, volume I, pages 379–382, Lausanne, September 16-19 1996. IEEE Sig Proc Society.

[23] E. P. Simoncelli and W. T. Freeman. The steerable pyramid: A flexible architecture for multi-scale derivative computation. In *IEEE Int'l. Conf. on Image Proc.*, volume 3, pages 444–447, 1995.

[24] J.-L. Starck, E. J. Candes, and D. L. Donoho. The curvelet transform for image denoising. *Image Processing, IEEE Transactions on*, 11(6):670–684, 2002.

[25] Y. Teh, M. Welling, and S. Osindero. Energy-based models for sparse overcomplete representations. *Journal of Machine Learning Research*, 4:1235–1260, 2003.

[26] B. Wegmann and C. Zetzsche. Statistical dependence between orientation filter outputs used in an human vision based image code. In *Proc Visual Comm. and Image Processing*, volume 1360, pages 909–922, Lausanne, Switzerland, 1990.

[27] P. Winkler. *Image Analysis, Random Fields And Markov Chain Monte Carlo Methods*. Springer, 2nd edition, 2003.

[28] S. C. Zhu, Y. N. Wu, and D. Mumford. Minimax entropy principle and its application to texture modeling. In *Neural Computation*, volume 9, pages 1627–1660, 1997.