

# Image Modeling and Denoising With Orientation-Adapted Gaussian Scale Mixtures

David K. Hammond and Eero P. Simoncelli, *Senior Member, IEEE*

**Abstract**—We develop a statistical model to describe the spatially varying behavior of local neighborhoods of coefficients in a multiscale image representation. Neighborhoods are modeled as samples of a multivariate Gaussian density that are modulated and rotated according to the values of two hidden random variables, thus allowing the model to adapt to the local amplitude and orientation of the signal. A third hidden variable selects between this oriented process and a nonoriented scale mixture of Gaussians process, thus providing adaptability to the local orientedness of the signal. Based on this model, we develop an optimal Bayesian least squares estimator for denoising images and show through simulations that the resulting method exhibits significant improvement over previously published results obtained with Gaussian scale mixtures.

**Index Terms**—Gaussian Scale Mixtures, image denoising, image processing, statistical image modeling, wavelet transforms.

## I. INTRODUCTION

THE set of natural photographic images is a distinct subset of the space of all possible 2-D signals, and understanding the distinguishing properties of this subset is of fundamental interest for many areas of image processing. For example, restoring images that have been corrupted with additive noise relies upon describing and exploiting the differences between the desired image signals and the noise process. Here, we characterize these differences statistically.

In a statistical framework, each individual photographic image is viewed as a sample from a random process. Since it is difficult to characterize such a process in a high-dimensional space, it is common to assume that the model should be translation-invariant (stationary). The most well-known example is that of power spectral models, which describe the Fourier coefficients of images as samples from independent Gaussians, with variance proportional to a power function of the frequency [1]. More recently, many authors have modeled the marginal statistics of images decomposed in multiscale bases using generalized Gaussian densities that exhibit sparse behavior (e.g., [2]–[4]). Denoising algorithms based on marginal models naturally take the form of 1-D functions (such as thresholding) that are applied pointwise and uniformly across the transform coefficient domain. However, natural images typically contain

diverse content, with high-contrast features such as edges, textured regions, and low-contrast smooth regions appearing in different spatial locations. Thus, a denoising calculation that is appropriate in one region may be inappropriate in another.

One means of constructing a model that is statistically homogeneous, but still able to adapt to spatially varying signal behaviors is by parametrizing a local model with variables that are themselves random. Such *doubly stochastic* or hierarchical statistical model have found use in a wide variety of fields, from speech processing to financial time series analysis. For image modeling, a number of authors have developed models with hidden variables that control the local amplitudes of multiscale coefficients [5]–[12]. These models can adapt to the spatially varying amplitude of local clusters of coefficients, a feature that clearly differentiates photographic images from noise. The special case in which clusters of coefficients are modeled as a product of a Gaussian random vector and a (hidden) continuous scaling variable is known as a *Gaussian scale mixture* (GSM) [13]. This model has been used as a basis for high-quality denoising results [14].

One of the most striking features that distinguishes natural images from noise is the presence of strongly oriented content. Although they account for the variability of local coefficient magnitudes, the models mentioned above do not explicitly capture the fact that many local image regions are dominated by a single orientation which varies with spatial location. Several authors have explored methods of adapting to the dominant local orientation [33], [34]. One method of introducing local adaptation to the GSM model by estimating covariances over local image regions has been recently studied in [15].

In this article, we extend the GSM model to describe patches of coefficients as Gaussian random variables, with a covariance matrix that depends on a set of three hidden variables associated with the signal amplitude, the “orientedness” (a measure of how strongly oriented the local signal content is), and the dominant orientation. The probability distribution over the set of all patches thus takes the form of a mixture of Gaussians that are parametrized by these hidden variables. We validate our model by using it to construct a Bayesian least squares estimator for removing additive white noise. The resulting denoising algorithm yields performance that shows significant improvement over the results obtained with the GSM model [14], and is close to the current state of the art. A brief and preliminary version of this work was described in [16].

## II. STATISTICAL MODELS

The use of multiscale oriented decompositions (loosely speaking, *wavelets*) has become common practice in image

Manuscript received December 21, 2007; revised July 28, 2008. Current version published October 10, 2008. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Stanley J. Reeves.

D. K. Hammond is with the Ecole Polytechnique Federale de Lausanne, 1015 Lausanne, Switzerland (e-mail: david.hammond@epfl.ch).

E. P. Simoncelli is with the Howard Hughes Medical Institute, Center for Neural Science, and the Courant Institute of Mathematical Sciences, New York University, New York, NY 10003 USA (e-mail: eero.simoncelli@nyu.edu).

Digital Object Identifier 10.1109/TIP.2008.2004796

processing. As natural image signals typically produce sparse responses in the wavelet domain, whereas noise processes do not, modeling the differences between image and noise signals becomes easier in the wavelet domain than in the original pixel domain. Additionally, constructing the model in the wavelet domain allows information at multiple spatial scales to be treated in a consistent manner. The models presented in this article are all models for local patches of wavelet coefficients.

### A. Steerable Pyramid Representation

In order to construct a model that can adapt to local orientation, we require a representation that allows measurement of local orientation and rotation of patches of subband coefficients. Critically sampled separable orthogonal wavelet representations are unsuitable for this: the filters in each subband do not span translation-invariant or rotation-invariant subspaces, and, thus, the coefficients cannot be interpolated at intermediate orientations or spatial positions because of the aliasing effects induced by critical sampling.

As an alternative, we use a *steerable pyramid* (SP) [17], an overcomplete linear transform that has basis functions comprised of oriented multiscale derivative operators. The steerable pyramid transform of height  $J$  decomposes an image into a highpass residual subband,  $J$  sets of oriented bandpass bands at dyadically subsampled scales, and a residual lowpass band. The full set of bandpass filters for the SP can be obtained from dilations, translations and rotations of a single “mother wavelet.” The SP transform may be constructed with an arbitrary number  $K$  of orientation subbands. When  $K = 2$  the bandpass filters are first order derivative operators in the  $x$  and  $y$  directions. In this case, the oriented subbands of the SP provide a measure of the image gradient at multiple scales.

Unlike separable wavelets, the SP filters centered at a single point span a rotation invariant subspace. As such, any of the filters may be rotated to any angle simply by taking a linear combination of the full set of filters, and this in turn means that the entire basis may be rotated to any desired orientation. Throughout this paper, we use the phrase “wavelet coefficients” to refer to the steerable pyramid coefficients, unless otherwise indicated.

### B. Local Coefficient Patches and the GSM Model

Models based on marginal statistics of wavelet coefficients are appealing due to their relative tractability, but make the implicit assumption that the coefficients are independent. But wavelet coefficients of natural images from neighboring space and scale locations show strong statistical interdependencies. These have been described and exploited by a number of authors for applications such as image coding, texture synthesis, block artifact removal and denoising [6], [18], [19]–[21]. These dependencies arise from the fact that images contain sparse localized features, which induce responses in clusters of filters that are nearby in space, scale, and orientation.

A natural means of capturing these dependencies is to construct multivariate probability models for small patches of wavelet coefficients. For denoising purposes, each coefficient can then be estimated based on a collection of coefficients centered around it [14]. Including a “parent” coefficient from a coarser scale subband, as shown in Fig. 1 allows the model

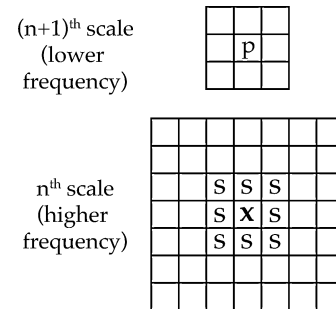


Fig. 1. Generalized neighborhood of wavelet coefficients. The diagram indicates an example generalized patch for a coefficient at position (x) consisting of sibling (s) and parent (p) coefficients.

to take advantage of cross-scale dependencies in a natural manner [20], [22]. Note that in this type of local model, we do not segment the image into nonoverlapping patches, since this would likely introduce block-boundary artifacts. We instead apply the local model independently on overlapping patches, and simply ignore the effects of the patch overlap.

Perhaps the most noticeable dependency among nearby wavelet coefficients is the clustering of large magnitude coefficients near image features: the presence of a single large coefficient indicates that other large coefficients are likely to occur nearby. This behavior is the primary inspiration for the Gaussian Scale Mixture (GSM) model [9]. Letting  $v \in \mathbb{R}^d$  denote a patch of  $d$  coefficients, the GSM model is

$$v = \sqrt{z}u \quad (1)$$

where  $u$  is a zero mean Gaussian with fixed covariance  $C$ , and  $z$  is the spatially varying scalar hidden variable.

The GSM model explains the inhomogeneity in the amplitudes of wavelet coefficients through the action of the scalar hidden variable,  $z$ , which modulates a homogeneous Gaussian process. One property of the model is that if the action of this scalar multiplier is “undone,” by dividing an ensemble of coefficient vectors by their hidden variables  $z$ , the subsequent ensemble of transformed vectors would have homogeneous Gaussian statistics. It has been observed that transforming filtered image data by dividing by the local standard deviation yields marginal statistics that are closer to Gaussian [23], [9].

This is illustrated in Fig. 2 for a single steerable pyramid subband. The original subband marginal statistics are far from Gaussian, as can be seen from examining the log histogram. We introduce the notation

$$g(v; C) = \frac{1}{\sqrt{(2\pi)^d |C|}} \exp\left(-\frac{1}{2}vC^{-1}v^T\right) \quad (2)$$

for the zero mean multivariate Gaussian with covariance  $C$ . Each patch  $x$  may be viewed as a sample of the Gaussian  $g(x; zC)$ , where  $C$  may be estimated for the entire band by taking the sample covariance of all of the extracted overlapping coefficient patches. For each individual patch  $x_i$ , the maximum likelihood estimate of  $z$  is given by

$$\hat{z}_i = \operatorname{argmax}_z g(x_i; zC) = \frac{1}{d} x_i^T C^{-1} x_i \quad (3)$$

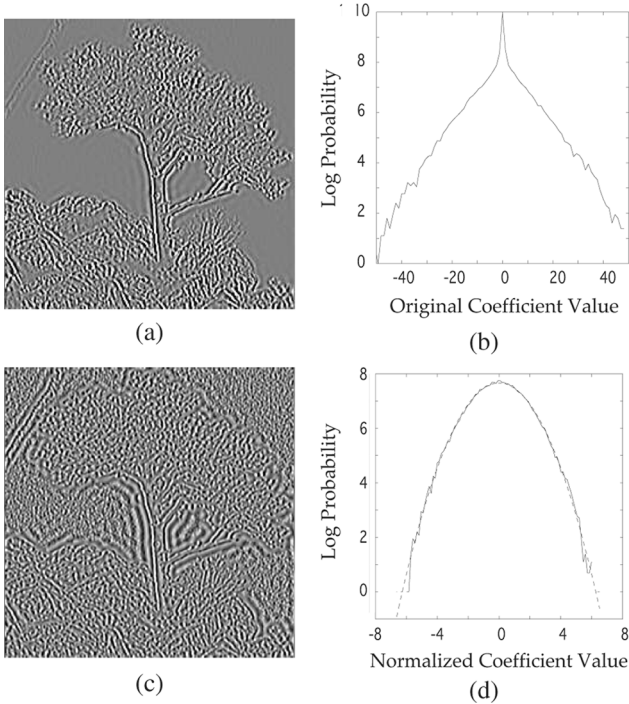


Fig. 2. Effects of divisive normalization. (a) Original subband; (b) log-histogram of marginal statistics for original subband; (c) subband normalized by estimated  $\hat{z}$  at each location; (d) log-histogram for normalized coefficients, showing Gaussian behavior. Dashed line is parabolic curve fit to the log histogram, for comparison.

Dividing each coefficient by the value of  $\sqrt{\hat{z}}$  computed using a neighborhood centered around it gives the transformed subband shown in Fig. 2(c). As can be seen visually, the local contrast of this normalized subband is much more spatially homogeneous than for the original subband. The marginal statistics are also much closer to Gaussian, as may be seen by examining the log-histogram, shown in Fig. 2(d).

### C. OAGSM Model

We extend the GSM model to an *orientation-adaptive GSM* (OAGSM) model by incorporating adaptation to local signal orientation using a second spatially varying hidden variable  $\theta$ . Patches under the OAGSM model may be viewed as having been produced through the following generative process. At each location the  $z$  and  $\theta$  variables are first drawn according to a fixed prior distribution. A coefficient patch  $v$  is formed by drawing a sample from a fixed multivariate Gaussian process, rotating it by  $\theta$  (around the center of the patch), and scaling by  $\sqrt{z}$ . This implies

$$v = \sqrt{z}R(\theta)u \quad (4)$$

where  $R(\theta)$  is an operator performing rotation about the center of the patch, and  $u$  is a zero mean multivariate Gaussian with fixed covariance  $C_0$ .

Rotating by  $R(\theta)$  and scaling by  $\sqrt{z}$  are both linear operations. This implies that when conditioned on fixed values for the hidden variables,  $v$  is simply a linearly transformed Gaussian and is, thus, itself Gaussian, with covariance

$$zC(\theta) = zR(\theta)C_0R(\theta)^T \quad (5)$$

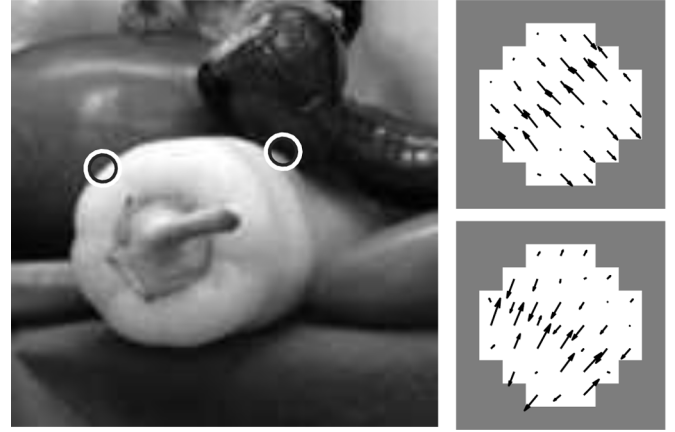


Fig. 3. Patches of two-band SP coefficients, displayed as fields of gradient vectors, taken from two oriented regions with different orientations. The structures of the two patches are similar, apart from the change in dominant orientation.

where the “oriented covariances”  $C(\theta)$  are defined as rotated versions of base covariance  $C_0$ .

The OAGSM model relies on the idea that differences in structure between coefficient patches in different oriented regions may be explained by the action of  $\theta$ . This is illustrated in Fig. 3, where the structure of two wavelet patches in different oriented regions is seen to be similar up to rotation. Attempting to describe the statistics of an ensemble of such patches without accounting for the rotational relationship between them will result in mixing structures at different orientations together. Such inappropriate data pooling will result in a less powerful signal description as some of the structure will have been averaged out. Conversely, taking advantage of this rotational relationship between coefficient patches can lead to a simpler and more homogeneous model.

This statement can be made more precise by analyzing the second-order covariance statistics for ensembles of coefficient patches. Analogous to the divisive normalization described in Section II-B, we can “undo” the action of the rotator hidden variable by estimating the dominant local orientation of each patch, and then rotating each patch around its center by the estimated orientation. Performing this “orientation normalization” on every patch of coefficient from a particular spatial scale gives an ensemble of transformed patches with the same dominant orientation. These rotated patches are more homogeneous, and, therefore, easier to describe compactly, than the ensemble of raw original patches.

One simple way to quantify this is to examine the energy compaction properties of the two representations by performing principle component analysis (PCA) on both sets of patches. Let  $v_i^{\text{raw}}$  and  $v_i^{\text{rot}}$  denote the raw and rotated “vectorized” patches extracted from one spatial scale of an image. We define sample covariance matrices for each ensemble

$$C^{\text{raw}} = \sum_i v_i^{\text{raw}} (v_i^{\text{raw}})^T$$

$$C^{\text{rot}} = \sum_i v_i^{\text{rot}} (v_i^{\text{rot}})^T$$

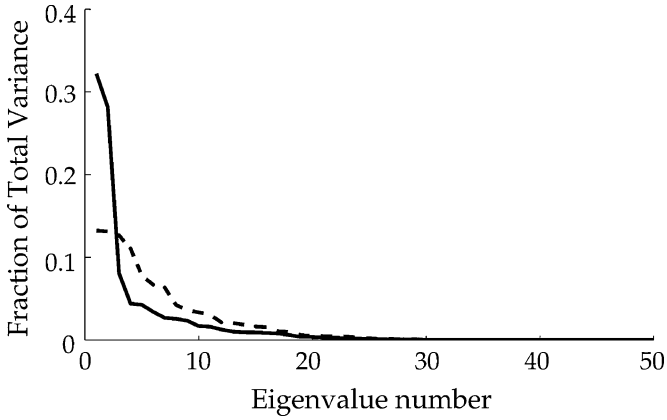


Fig. 4. Normalized eigenvalues of covariance matrix estimated from coefficient patches drawn from single scale of the pyramid representation of an example image (“peppers”). Dashed curve corresponds to raw patches, and solid curve to patches rotated according to dominant orientation.

and then examine their eigenvalues and eigenvectors. If the eigenvalues are normalized by the trace of the covariance, then they may be interpreted as the fraction of total signal variance that lies along the direction of each corresponding eigenvector. Normalized eigenvalues for  $C^{\text{rot}}$  and  $C^{\text{raw}}$  are plotted in decreasing order in Fig. 4. Comparing the results for the raw and rotated patches, we see that a greater portion of the total variance in the rotated patches is accounted for by a smaller number of dimensions. This “energy compaction” should translate into an advantage for removing additive white noise, since it amplifies the signal relative to the noise.

#### D. OAGSM/NC Model

Although orientation is an important property of natural image patches, images may also contain features such as corners and textures that are either unoriented or of mixed orientation. Modeling such nonoriented areas with the OAGSM process may lead to inappropriate behavior, such as the introduction of oriented artifacts during denoising. To better model such regions, we augment the OAGSM model by mixing it with a nonoriented adaptive process that is a simple GSM. Selection between the oriented and nonoriented processes is controlled by a third spatially varying binary hidden variable  $\delta$ , which allows adaptation to local signal “orientedness.” The generative process for the this Orientation-adapted GSM with nonoriented component (OAGSM/NC) model is given by

$$v = \begin{cases} \sqrt{z}R(\theta)u^{\text{ori}}, & \text{if } \delta = 1 \\ \sqrt{z}u^{\text{nor}}, & \text{if } \delta = 0 \end{cases} \quad (6)$$

where  $u^{\text{ori}}$  is a sample from a Gaussian with covariance  $C_0$  and  $u^{\text{nor}}$  is a sample from a Gaussian with covariance  $C_{\text{nor}}$ . As before for the OAGSM model, samples from the distribution described by  $C_0$  should represent oriented structure at a fixed nominal orientation. Intuitively, the nonoriented component determined by  $C_{\text{nor}}$  is used to describe the “leftovers,” i.e., image regions not well captured by the oriented component of the model.

It follows from (6) that the probability distribution for  $v$  when conditioned on the hidden variables is

$$p(v|z, \theta, \delta) = g(v; \delta z C(\theta) + (1 - \delta) z C_{\text{nor}}). \quad (7)$$

The complete density on the patch  $v$  is then computed by integrating against the prior density of the hidden variables

$$p(v) = \int p(v|z, \theta, \delta) p(z, \theta, \delta) dz d\theta d\delta. \quad (8)$$

We assume a separable prior density for the hidden variables,  $p(z, \theta, \delta) = p(z)p(\theta)p(\delta)$ . Following [14], the prior on  $z$  is derived from the so-called Jeffrey’s noninformative pseudo-prior  $p(z) \propto (1/z)$  [10]. This “density” cannot be normalized unless  $z$  is truncated within some range  $[z_{\text{min}}, z_{\text{max}}]$ . The Jeffrey’s pseudo-prior is equivalent to placing a uniform density on  $\log z$ .

For the  $\theta$  hidden variable we use a constant prior. The oriented covariances used in this work are  $\pi$ -periodic, i.e.,  $C(\theta) = C(\theta + \pi)$ . This is a consequence of the Steerable Pyramid filters being either symmetric or anti-symmetric with respect to rotation by  $\pi$ . Accordingly, we choose a prior for  $\theta$  that is constant on the interval  $[0, \pi]$ . This is equivalent to assuming rotation invariance of natural signals, i.e., that oriented content is equally likely to occur at any orientation. While this assumption may not hold exactly for classes of images with strong preference for particular orientations, such as landscapes or city scenes taken with the camera parallel to the horizon, it is a natural generic choice. If desired, the prior could be modified to incorporate higher probabilities for orientations such as vertical or horizontal.

Finally,  $\delta$  is a binary variable, and its density is simply a two component discrete density. Fixing our notation, we set  $\beta$  to be  $p(\delta = 1)$ , i.e., the prior probability of drawing each patch from the oriented process. Note that integrating over  $\delta$  in (8) implies

$$\begin{aligned} p(v) &= \beta \int g(v; zC(\theta)) p(z) p(\theta) dz d\theta \\ &\quad + (1 - \beta) \int g(v; zC_{\text{nor}}) p(z) dz \\ &= \beta p_{\text{ori}}(v) + (1 - \beta) p_{\text{nor}}(v). \end{aligned} \quad (9)$$

This expression makes it clear that as  $\beta$  varies between 0 and 1, the OAGSM/NC model interpolates between the oriented and nonoriented component models. Unlike the priors for  $z$  and  $\theta$  which we assume are fixed and not data adaptive, we treat  $\beta$  as a parameter to be estimated for each wavelet subband. This allows the model to handle image subbands that have different proportions of oriented versus nonoriented content.

### III. ESTIMATING MODEL PARAMETERS

The parameters for the full OAGSM/NC model consist of the oriented covariances  $C(\theta)$ , the nonoriented covariance  $C_{\text{nor}}$  and the orientedness prior  $\beta$ . We can fit these parameters to the noisy data for each subband of the SP transform of the image to be denoised.

#### A. Oriented Covariances

The oriented covariances  $C(\theta)$  for fixed  $\theta$  are defined by the expectation  $E[R(\theta)u(R(\theta)u)^T]$ . We compute these by a “patch rotation” method analogous to that used in the example of Fig. 4. At each spatial location, the dominant local orientation is estimated. Covariances are then computed by spatially rotating each of the patches in the subband, and calculating the sample outer product of this rotated ensemble.

In principle, the oriented covariances could also be computed by measuring  $C_0 = E[ww^T]$  and applying (5). However, there is a technical difficulty with defining the rotation operator  $R(\theta)$  for patches of coefficients that are not embedded within a larger subband. Rotating a square patch of coefficients will require access to coefficients outside of the original square, to compute values for the “corners” of the rotated patch which will arise from image signal outside of the original square. This implies that the domain of  $R(\theta)$  must be larger than its range. In this paper we avoid this difficulty by only applying  $R(\theta)$  to patches “in context,” i.e., that are embedded within a complete subband of coefficients.

The OAGSM model describes an ensemble of noisy oriented wavelet patches  $\{w_i\}$ , generated by local hidden variables  $z_i$  and  $\phi_i$ , as

$$w_i = \sqrt{z_i}R(\phi_i)u_i + n_i \tag{10}$$

where  $n_i$  are samples of the noise process. The primary impediment to computing the oriented covariances is that the hidden rotator variables  $\phi_i$  are different for each patch. A method for resolving this difficulty is suggested by the following thought experiment. Suppose one had access to an ensemble of noisy patches  $w_i$  that were formed from a *single* fixed value of the hidden variable, i.e.,  $\phi_i = \theta^*$  for all  $i$ . Taking the sample outer product of this ensemble (suppressing the index  $i$ ) would give

$$\begin{aligned} E[ww^T] &= E[zR(\theta^*)ww^TR(\theta^*)] + E[nn^T] \\ &= E[z]R(\theta^*)E[ww^T]R(\theta^*) + C_n \tag{11} \\ &= E[z]C(\theta^*) + C_n. \tag{12} \end{aligned}$$

Note that we have assumed that the noise and signal are independent. The oriented covariance for angle  $\theta^*$  may thus be computed by taking the sample outer product of the ensemble, subtracting off the noise covariance  $C_n$ , and dividing by the expected value of  $z$ . This is feasible, since we assume both the noise process and the distribution of  $z$  are known. Note that as the oriented covariance estimates are computed from data, it is possible that they are no longer positive definite after subtracting  $C_n$ . In practice, we impose positive definiteness by diagonalizing the estimated covariances and replacing all negative eigenvalues by a small positive constant.

A collection of patches drawn from a fixed value  $\theta^*$  of the rotator hidden variable is not immediately accessible. Instead, we produce such an ensemble by manipulating the patches present in the given noisy image. This manipulation relies on the idea that rotating an image region around its center will change only the orientation of the region, while preserving the rest of its structure. Given a coefficient patch modeled as a sample of an OAGSM with a particular value  $\phi$  for the rotation hidden variable, rotating the underlying image signal by  $\Delta\phi$  and recomputing the filter responses will give a patch that is equivalent to an OAGSM sample with hidden variable  $\phi + \Delta\phi$ . The true values of the rotator hidden variables  $\phi_i$  corresponding to the given ensemble of raw patches are unknown, but can be estimated by computing the dominant neighborhood orientation  $\phi_i^*$  of the noisy patches (see next subsection).

Given a set of noisy patches  $w_i$  with estimated neighborhood orientations  $\phi_i^*$ , we rotate each patch by  $\theta - \phi_i^*$  to produce an ensemble of patches that is approximately equivalent to one produced by the OAGSM process with a fixed value  $\theta$  for the rotator variable. Taking the sample outer product of this rotated ensemble then yields, following (12)

$$E[(R(\theta - \phi_i^*)w_i)(R(\theta - \phi_i^*)w_i)^T] = E[z]C(\theta) + C_n \tag{13}$$

from which  $C(\theta)$  can be calculated. This computation must be repeated for each value of  $\theta$  for which  $C(\theta)$  is required. In practice, we sample  $\theta$  at a relatively small number of values.

1) *Estimating Neighborhood Orientation:* Computing the dominant orientation at each point in space requires a measure of the image gradient. We use the SP with two orientation bands to estimate the local orientations. Note that this does not restrict the order of the SP transform used for denoising, as the estimated orientations may be used to rotate coefficient patches for a SP with a different number of orientation subbands. Consider an  $m \times m$  patch of two-band SP coefficients  $v$  as a collection of  $m^2$  gradient vectors  $h_i$  for  $i = 1 \dots m^2$ . We define the neighborhood orientation  $\phi$  for the patch as the angle of the unit vector  $k(\phi) = (\cos(\phi), \sin(\phi))^T$  that maximizes the sum of squares of inner products

$$\sum_{i=1}^{m^2} (k(\phi)^T h_i)^2. \tag{14}$$

This is equivalent to the orientation of the eigenvector corresponding to the largest eigenvalue of the  $2 \times 2$  *orientation response matrix*  $M = \sum h_i h_i^T$ . Writing  $h_i = (a_i, b_i)^T$ , the dominant orientation is given explicitly by

$$\phi = \frac{1}{2} \angle (2\sum a_i b_i, \sum (a_i^2 - b_i^2)) \tag{15}$$

where  $\angle$  indicates the angle of the vector whose components are specified by the two arguments.

2) *Patch Rotation:* We describe the precise form of the patch rotation operator  $R(\theta)$  first mentioned in (4). Rotating a patch of wavelet coefficients is equivalent to finding the coefficients that would arise if the underlying image signal was rotated around the center of the patch. In order to write this precisely, introduce the following notation. Let  $f(x, y)$  be the original continuous image signal. Assume that the number of orientation bands for the steerable pyramid transform being used is fixed. Let  $B_\phi^s(x, y)$  denote the SP filter in the space domain centered at the origin, with orientation  $\phi$  at scale  $s$ .

Let  $i = 1 \dots d$  denote the different coefficients of the wavelet patch. Each coefficient corresponds to a SP filter at a particular location, orientation and scale. Let  $(x_i, y_i)$  denote the position of the filter for the  $i$ th coefficient, relative to the center of the patch. In this work, the filters corresponding to a given patch have the same orientation  $\phi$ . As we are considering patches that include parent coefficients, however, the spatial scale of the filter will depend on the patch index  $i$ .

Let  $v$  be the wavelet patch centered at the origin. We then have

$$v_i = \int \int f(x, y) B_\phi^{s_i}(x - x_i, y - y_i) dx dy. \tag{16}$$

We are now prepared to calculate  $R(\theta)v$ . By the definition of patch rotation, we have

$$(R(\theta)v)_i = \int \int f(r_{-\theta}(x, y))B_{\phi}^{s_i}(x - x_i, y - y_i)dx dy \quad (17)$$

where  $r_{\theta}(x, y) = (x \cos \theta + y \sin \theta, -x \sin \theta + y \cos \theta)$ . This integral will be invariant under the change of variables given by rotating the coordinate axes by  $\theta$ . This yields

$$(R(\theta)v)_i = \int \int f(x, y)B_{\phi-\theta}^{s_i}(x - x'_i, y - y'_i)dx dy \quad (18)$$

where  $(x'_i, y'_i) = r_{\theta}(x_i, y_i)$ . Thus, the  $i^{\text{th}}$  coefficient of the transformed patch is given by the response of the original, unrotated image to a filter with orientation  $\phi - \theta$  at location  $(x'_i, y'_i)$ .

Performing patch rotation thus requires computing the response of rotated filters that may not lie on the original sample lattice. These may be computed from the original transform coefficients using the steerability and shift invariance properties of the steerable pyramid. Steerability implies that a filter rotated to any orientation can be decomposed as a sum of  $K$  filters at the “standard orientations”  $\phi_k = ((k - 1)\pi/K)$  for  $k = 1 \dots K$ . This implies that there exist “steering functions”  $c_k(\cdot)$  such that

$$B_{\phi-\theta}^s = \sum_{k=1}^K c_k(\phi - \theta)B_{\phi_k}^s. \quad (19)$$

Details of calculating these  $c_k(\cdot)$  may be found in [17] and [24].

The shift invariance property of the transform follows from the design of the SP filter responses, which decay smoothly to zero for frequencies approaching  $\pi$ . This implies that there is no aliasing in each subband, so that the responses for each subband to any shifted signal may be computed by interpolating that subband directly. This should be contrasted with the behavior of critically sampled orthogonal wavelet transforms, which have severe aliasing which prevents spatial interpolation off of the sample lattice.

In this work, interpolation is done by upsampling by a factor of  $2\gamma$  in each direction (by zero padding the Fourier transform of each subband and inverting), followed by bilinear interpolation. All calculations in this paper were performed using  $\gamma = 3$ . Using this shift invariance property, the responses of the “standard orientation” filters can be first interpolated at locations  $(x'_i, y'_i)$ , off of the original sample lattice, and then transformed using (19) to give the filter responses corresponding to each element  $(R(\theta)v)_i$  of the rotated patch.

### B. OAGSM/NC Parameters

The remaining model parameters for the full OAGSM/NC model are the nonoriented covariance  $C_{\text{nor}}$  and the orientedness prior  $\beta$ .  $C_{\text{nor}}$  is computed as in [14], by taking the average outer product of the raw, nonrotated coefficient patches. This is essentially the same procedure as described in the previous section, but without applying patch rotation.

We compute  $\beta$  by maximizing the likelihood of the model. The complete OAGSM/NC model can be written as

$$p(v) = (1 - \beta)p_{\text{nor}}(v) + \beta p_{\text{ori}}(v) \quad (20)$$

where, in the presence of noise, we have

$$p_{\text{nor}}(v) = \int g(v; zC_{\text{nor}} + C_n)p(z)dz$$

$$p_{\text{ori}}(v) = \int g(v; zC(\theta) + C_n)p(z)p(\theta)dz d\theta. \quad (21)$$

Given a collection of  $m$  patches  $v^k$ , typically all of the patches from a particular subband, the log-likelihood function will be

$$L(\beta) = \sum_{k=1}^m \log((1 - \beta)p_{\text{nor}}(v^k) + \beta p_{\text{ori}}(v^k)). \quad (22)$$

Directly maximizing this function over  $\beta$  is problematic, because of the sum present inside of the logarithm. Instead, we employ the Expectation Maximization (EM) algorithm, a widely used iterative method for performing maximum likelihood estimation [25]. The E-step consists of computing the expectation, by integrating over the so-called “missing data,” of the “full data” likelihood given the previous iterate of the parameter estimates. The resulting average likelihood function no longer contains any reference to the hidden variables, but is a function of the parameters to be fit. The M-step consists of computing the arg-max of this function, which gives the next iterates of the estimated parameters.

The following brief explanation of EM for estimating the component weights of a mixture model follows the treatment in [26]. For this problem, the “missing data” consists of binary indicator variables  $\tau_k^\delta$ , where  $k = 1 \dots m$  and  $\delta = 0, 1$ . For each value of  $k$ , exactly one of these variables will equal 1, indicating which component the  $k^{\text{th}}$  sample arose from, e.g., we have  $\tau_k^0 = 1$  and  $\tau_k^1 = 0$  if the  $k^{\text{th}}$  sample arose from the nonoriented component, or  $\tau_k^0 = 0$  and  $\tau_k^1 = 1$  if the  $k^{\text{th}}$  sample arose from the oriented component.

Let  $\tau_k = (\tau_k^0, \tau_k^1)$ , and let  $\vec{\tau}$  indicate the entire set of indicator variables. The complete-data likelihood will be a product of the terms  $p(v^k, \tau_k | \beta) = p(v^k | \tau_k, \beta)p(\tau_k | \beta)$ . As  $\tau_k$  indicates which component  $v^k$  arose from, we have

$$p(v^k | \vec{\tau}, \beta) = (p_{\text{nor}}(v^k))^{\tau_k^0} (p_{\text{ori}}(v^k))^{\tau_k^1}$$

$$p(\tau_k | \beta) = (1 - \beta)^{\tau_k^0} \beta^{\tau_k^1}. \quad (23)$$

The complete data log-likelihood  $\log(p(V, \tau | \beta))$  may then be written as

$$\sum_k \tau_k^0 \log(p_{\text{nor}}(v^k)) + \tau_k^1 \log(p_{\text{ori}}(v^k))$$

$$+ \tau_k^0 \log(1 - \beta) + \tau_k^1 \log(\beta). \quad (24)$$

As the “missing data” variables  $\tau$  appear linearly in the above expression, the expectation operation for the E step can be passed into the above sum. The E step may

thus be done by replacing each occurrence of  $\tau_k^\delta$  by  $t_k^\delta(\beta) = E[\tau_k^\delta | \beta, v^k] = p(\tau_k^\delta = 1 | \beta, v^k)$ . By Bayes' rule this is  $(p(v^k | \tau_k^\delta = 1, \beta) p(\tau_k^\delta = 1 | \beta) / p(v^k | \beta))$ . Evaluating for  $\delta = 0, 1$  shows

$$\begin{aligned} t_k^0(\beta) &= \frac{p_{\text{nor}}(v^k)(1-\beta)}{(1-\beta)p_{\text{nor}}(v^k) + \beta p_{\text{ori}}(v^k)} \\ t_k^1(\beta) &= \frac{p_{\text{ori}}(v^k)\beta}{(1-\beta)p_{\text{nor}}(v^k) + \beta p_{\text{ori}}(v^k)}. \end{aligned} \quad (25)$$

The first two terms of (24) do not contain  $\beta$  and may be ignored. The M step to find the  $n$ th iterate for  $\beta$  consists of finding the maximum of

$$Q(\beta, \beta^{n-1}) = \sum_{k=1}^m t_k^0(\beta^{n-1}) \log(1-\beta) + t_k^1(\beta^{n-1}) \log(\beta). \quad (26)$$

A simple calculation yields the maximum at

$$\beta^n = \frac{1}{m} \sum_k t_k^1(\beta^{n-1}). \quad (27)$$

Thus, at every step,  $\beta$  is replaced by the average expected probability that each sample arose from the oriented component, conditioned on the previous iterate of  $\beta$ .

In general, the EM algorithm is only guaranteed to converge to a local maximum of the likelihood function. For this problem, however,  $L(\beta)$  is concave down as its second derivative

$$\frac{d^2 L}{d\beta^2} = \sum_k \frac{-(p_{\text{ori}}(v^k) - p_{\text{nor}}(v^k))^2}{((1-\beta)p_{\text{nor}}(v^k) + \beta p_{\text{ori}}(v^k))^2} \quad (28)$$

is negative for all  $\beta \in (0, 1)$ . Thus,  $L$  can have only a single maximum, and the EM procedure will converge to the ML estimate for  $\beta$ . In practice for the OAGSM/NC model, roughly 20 iterations are required for reasonable convergence.

#### IV. DENOISING ALGORITHM

We study the problem of removing additive Gaussian noise from photographic images as both a validation of the OAGSM/NC model, as well as a useful practical application. We follow a Bayesian formalism, where the OAGSM/NC model is used as a prior distribution for the statistics of clean signal transform coefficients. All denoising calculations are performed in the space of steerable pyramid coefficients.

##### A. Bayesian Estimator

We model a generalized patch of noisy wavelet coefficients  $y \in \mathbb{R}^d$  as  $y = x + n$  where  $x$  is the original signal and  $n$  is the additive Gaussian noise. Let  $C_n$  be the  $d \times d$  covariance matrix of the noise process in the wavelet domain. Note that as the SP transform is not orthogonal, the noise process in each subband will be correlated even if it is white in the pixel domain.

In the Bayesian framework, both the desired signal  $x$  and the noise process  $n$  are considered to be random variables, and denoising is performed by statistical estimation. For a particular noisy observation  $y$ , there are infinitely many possible signals

$x$  that may have produced the given observation. These candidates are not all equally likely, however, and follow the posterior probability distribution  $p(x|y)$ . One may view the estimating function  $\hat{x}(y)$  as selecting an element of this set according to some criterion. One common criterion is to choose  $\hat{x}(y)$  minimizing the expected squared error  $E = \int p(x|y) \|\hat{x}(y) - x\|^2 dx$ . The minimum error is achieved for  $\hat{x}(y) = \int xp(x|y)dx$ , the *a posteriori* mean. This Bayesian Minimum Mean Square Error (MMSE) estimator is used for the denoising calculations in the current work.

Both the OAGSM and the OAGSM/NC models described in this paper, as well as the original GSM model, consist of mixtures of Gaussian components parametrized by a set of hidden variables. For models of this form, an exact form for the Bayesian MMSE estimator can be calculated as an integral over the hidden variables. Let  $\vec{\kappa}$  denote the hidden variables, where then  $\vec{\kappa} = z$  for the original GSM,  $\vec{\kappa} = (z, \theta)$  for the OAGSM and  $\vec{\kappa} = (z, \theta, \delta)$  for the OAGSM/NC model. By conditioning on the hidden variables the posterior probability can be written as  $p(x|y) = \int p(x|\vec{\kappa}, y) p_{\vec{\kappa}}(\vec{\kappa}|y) d\vec{\kappa}$ . Inserting this into the expression for the MMSE estimator and exchanging the order of integration yields

$$\hat{x}(y) = \int x \left( \int p(x|y, \vec{\kappa}) p(\vec{\kappa}|y) d\vec{\kappa} \right) dx \quad (29)$$

$$= \int p(\vec{\kappa}|y) \left( \int xp(x|y, \vec{\kappa}) dx \right) d\vec{\kappa}. \quad (30)$$

A key point is that the signal description is Gaussian when conditioned on the hidden variables. This implies that the inner integral on the r.h.s. of (30) is precisely the expression for the MMSE of the Gaussian signal with covariance  $C(\vec{\kappa})$  corrupted by the Gaussian noise process  $n$ . This is a well known problem, with a linear (Wiener) solution  $W_{\vec{\kappa}}y = C(\vec{\kappa})(C_n + C(\vec{\kappa}))^{-1}y$ . We thus have

$$\hat{x}(y) = \int p(\vec{\kappa}|y) (W_{\vec{\kappa}}y) d\vec{\kappa}. \quad (31)$$

This is a weighted average of different Wiener estimates, where the weighting is controlled by  $p(\vec{\kappa}|y)$ . It is this weighting which allows the denoising algorithm to adapt to different local signal conditions. For noisy signal patches that are best described with power  $z^*$ , orientation  $\theta^*$  and orientedness  $\delta^*$ , the weights  $p(z, \theta, \delta|y)$  will be larger for values  $(z, \theta, \delta)$  closer to  $(z^*, \theta^*, \delta^*)$ , and smaller otherwise. The Wiener estimates  $W_{\vec{\kappa}}y$  will be more accurate when the hidden variables are closer to those that best describe the signal. As a result, the full estimate  $\hat{x}(y)$  will contain more contribution from the Wiener estimates that are more appropriate for the current noisy signal.

Computing the weightings  $p(\vec{\kappa}|y)$  is straightforward. Applying Bayes theorem gives

$$p(\vec{\kappa}|y) = \frac{p(y|\vec{\kappa})p(\vec{\kappa})}{p(y)} = \frac{p(y|\vec{\kappa})p(\vec{\kappa})}{\int p(y|\vec{\kappa})p(\vec{\kappa})d\vec{\kappa}}. \quad (32)$$

When conditioned on  $\vec{\kappa}$ ,  $y$  is the sum of the Gaussian signal with covariance  $C(\vec{\kappa})$  and the noise process. As the signal and noise are assumed independent, their sum is a Gaussian with

TABLE I  
 DENOISING PERFORMANCE FOR SIX IMAGES WITH FOUR DIFFERENT NOISE LEVELS. RESULTS ARE SHOWN FOR ALGORITHMS USING TWO ORIENTED SP BANDS (TOP HALF) EIGHT ORIENTED SP BANDS (BOTTOM HALF). VALUES GIVEN ARE AVERAGED OVER FIVE TRIALS. IN EACH CELL, NUMBER ON LEFT IS FOR THE OAGSM/NC ALGORITHM, NUMBER ON RIGHT IS FOR THE GSM ALGORITHM [14]. ALL VALUES INDICATE PSNR, COMPUTED AS  $20 \log_{10}(25.5/\sigma_e)$ , WHERE  $\sigma_e^2$  IS THE ERROR VARIANCE

$\sigma$ (PSNR)	Lena	Barbara	Boats	House	Peppers
10 (28.13)	35.52 / 35.21	34.00 / 33.39	33.48 / 33.34	35.27 / 34.94	33.70 / 33.29
20 (22.11)	32.57 / 32.15	30.29 / 29.54	30.24 / 30.08	32.29 / 31.89	30.26 / 29.79
25 (20.17)	31.58 / 31.16	29.07 / 28.36	29.21 / 29.06	31.28 / 30.89	29.16 / 28.71
50 (14.15)	28.46 / 28.07	25.29 / 24.93	26.17 / 26.05	28.17 / 27.82	25.76 / 25.46
75 (10.63)	26.67 / 26.37	23.48 / 23.27	24.56 / 24.49	26.34 / 26.02	23.81 / 23.60
10 (28.13)	35.67 / 35.51	34.26 / 33.96	33.56 / 33.49	35.43 / 35.27	33.78 / 33.54
20 (22.11)	32.78 / 32.55	30.68 / 30.27	30.37 / 30.29	32.48 / 32.25	30.39 / 30.06
25 (20.17)	31.81 / 31.58	29.51 / 29.11	29.36 / 29.29	31.50 / 31.25	29.30 / 28.97
50 (14.15)	28.70 / 28.48	25.92 / 25.67	26.34 / 26.29	28.33 / 28.10	25.95 / 25.71
75 (10.63)	26.90 / 26.72	23.92 / 23.81	24.74 / 24.71	26.52 / 26.29	24.03 / 23.88

covariance  $C(\vec{\kappa}) + C_n$ . This implies  $p(y|\vec{\kappa}) = g(y; C(\vec{\kappa}) + C_n)$ . The term  $p(\vec{\kappa})$  is the prior probability of the hidden variables, as described previously in Section II-D. Substituting these into (31) gives the full Bayes MMSE estimator

$$\hat{x}(y) = \frac{1}{Z} \int (W_{\vec{\kappa}} y) g(y; C_n + C(\vec{\kappa})) p(\vec{\kappa}) d\vec{\kappa} \quad (33)$$

where  $Z = \int g(y; C_n + C(\vec{\kappa})) p(\vec{\kappa}) d\vec{\kappa}$ .

In practice, we discretize the continuous hidden variables  $z$  and  $\theta$  and approximate the integral in (33) using a finite sum. For the results presented in this work, we used 13 sample values for  $z$  and 16 sample values for  $\theta$ . As mentioned in Section II-D, the Jeffrey's pseudo-prior  $p(z) \propto (1/z)$  used for  $z$  is formally equivalent to placing a uniform density on  $\log z$ . As  $z$  is sampled finitely, this is implemented by choosing sample values  $z_n$  uniformly logarithmically spaced between  $z_{\min}$  and  $z_{\max}$ . The results presented here employ  $\log(z_{\min}) = -20.5$  and  $\log(z_{\max}) = 3.5$ , the same parameters used in [14].

### B. Implementation Details

Noisy images are generated by adding synthetic white Gaussian noise to an original natural image. These are then decomposed with the steerable pyramid transform with a specified number of orientation subbands,  $K$ . We handle the image boundary by first extending the image by mirror reflection by 20 pixels in each direction, and then performing all convolutions for the steerable pyramid transform using circular boundary handling. After processing, this boundary segment is removed. Generalized patches are then extracted at each spatial scale. We used SP coefficient patches of  $5 \times 5$  siblings plus a parent coefficient (see Fig. 1), which were found to give the best denoising performance. From these noisy coefficients, both the oriented and nonoriented covariances were calculated as described in Section III. Recall that a two-band SP is used to measure the dominant local orientation that is used for rotating patches that are then used to compute oriented covariances. The denoising calculations are performed on a  $K$ -band SP (note, if  $K \neq 2$  then two SP transforms must be computed). The patches are then denoised at each scale using the MMSE estimator described above. This estimator  $\hat{x}(y)$  produces an estimate of the entire generalized patch. One could partition the transform domain into nonoverlapping square patches, denoise them separately, then invert the transform. However, doing this

would introduce block boundary artifacts in each subband. An alternative approach, used both here and in [14], is to apply the estimator to overlapping patches and use only the center coefficient of each estimate. In this way each coefficient is estimated using a generalized neighborhood centered on it.

The highpass and lowpass residual scalar subbands are treated differently. As in [14], we have used a modified steerable pyramid transform with the highpass residual band split into oriented subbands. This modified transform gives a denoising gain of about 0.2 dB over using the standard transform with a single (nonoriented) highpass residual. However, even subdivided into orientation subbands, the highpass filters are not steerable. This makes it difficult to obtain oriented covariances by the rotation method described in this paper. Accordingly, the highpass bands are denoised with the GSM model, without using the orientation or orientedness hidden variables.

The lowpass band typically has the highest signal-to-noise ratio. This follows as the power spectrum of natural images typically display a power-law decay with frequency, while the white noise process has constant power at all frequencies [23]. Additionally, for coarser spatial scales there are fewer available signal patches from which to fit the model parameters. At some point, the induced error from incorrectly estimating model parameters may become comparable to the benefit of denoising itself. This suggests that there is an effective limit to the number of spatial scales one can denoise. For this work, the pyramid representation is built to a depth of five spatial scales, with no denoising done for the lowpass band. After the estimation is done for the highpass and bandpass bands, the entire transform is inverted to give the denoised image.

### C. Results

We computed simulated denoising results on a collection of five standard 8-bit greyscale test images [14], four of size  $512 \times 512$  pixels and one of size  $256 \times 256$  pixels. In order to quantify the advantages of adaptation to orientation and orientedness, we provide comparisons against the results of denoising with the GSM model of [14]. In order to maintain consistency with the OAGSM/NC results, the GSM results presented here use a  $5 \times 5$  (plus parent) patch, as opposed to the  $3 \times 3$  (plus parent) patch used in [14]. We examined performance for steerable pyramids with both  $K = 2$  and  $K = 8$  oriented bands. We examined performance for five



different noise levels. Numerical results, presented as peak signal-to-noise ratio (PSNR), averaged over five realizations of the noise process, are presented in Table I.

The table shows a consistent improvement of OAGSM/NC over GSM, typically between 0.1 and 0.6 dB. The OAGSM/NC improvements are largest for images that have significant local orientation content, such as the “Barbara” image (which contains fine oriented texture in many regions), or the “Peppers” image. For images that have significantly more nonoriented texture, such as the Boats image, the improvement is often much smaller. Note also that the OAGSM/NC method offers a more substantial improvement over the GSM method for the 2-band representation.

The OAGSM/NC method also shows substantial visual improvement. Not surprisingly, this is most noticeable along strongly oriented image features. Details of two denoised images, with noise standard deviation  $\sigma = 25$ , are shown in Figs. 5 and 6. For the boat image, one can compare the appearance of the oblique mast. The contours of this object are clearly better preserved for the OAGSM/NC method than for the GSM method. The 2-band GSM denoised image shows clear artifacts due to the horizontal and vertical orientations of the underlying filters. Note that the OAGSM/NC method is able to significantly reduce these by adapting to local orientation, even though the estimation is performed using the same set of filters. The visual differences between the 8-band GSM and the 8-band OAGSM/NC methods are more subtle; however, the OAGSM/NC method suffers from less ringing along isolated edges such as the boat masts. Additionally, the appearance of the oriented texture on the shawl of the barbara image is more clearly preserved with the OAGSM/NC model.

The improvement in denoising performance afforded by the adaptation to orientation, while significant, may seem modest in light of the added complexity and computational cost of the full OAGSM/NC method. One explanation of this observation is that applying the standard GSM algorithm using filters that are highly tuned in orientation implicitly performs a type of adaptation to local orientation. For such a bank of oriented filters, the GSM covariance for a particular orientation subband will be computed by averaging over patches of filter responses from areas with different local orientations. However, the contribution from areas with local orientation significantly different from the orientation of the filters will be attenuated, due to the orientation tuning of the filters. Similarly, during denoising, image content in strongly oriented regions will be represented mostly by a few oriented subbands with similar orientation. The GSM denoising estimator will then mostly use these covariances for the denoising calculations, which provides implicit adaptation to the local signal orientation. The orientation tuning of the Steerable Pyramid filters becomes tighter with increasing number of orientation bands. This line of reasoning helps explain why the performance gains for the OAGSM/NC over the GSM model are more dramatic for the 2-band case than for the 8-band case.

The OAGSM/NC method introduces two hidden variables, allowing adaptation to both orientation and orientedness. In order to determine the importance of the adaptation to orientedness, we compared the performance of the OAGSM denoising al-



Fig. 5.  $128 \times 128$  detail from “Barbara” image, for noise with  $\sigma = 25$ . Top: Original, noisy (20.17 dB). Middle: gsm2 (28.36 dB), oagsmnc2 (29.07 dB). Bottom: gsm8 (29.11 dB), oagsmnc8 (29.51 dB).

TABLE II  
PSNR PERFORMANCE PENALTY INCURRED BY REMOVING ADAPTATION TO ORIENTEDNESS, FOR 2 SP BANDS (TOP) AND FOR 8 SP BANDS (BOTTOM)

$\sigma$ (PSNR)	Lena	Barbara	Boats	House	Peppers
25	0.085	0.019	0.085	0.065	0.009
50	0.109	-0.044	0.145	0.097	0.003
25	0.017	0.019	0.071	-0.005	-0.028
50	0.066	0.008	0.135	0.045	-0.025

gorithm without the nonoriented component to that of the full OAGSM/NC model. These differences are shown in Table II. In the 8-band case, allowing adaptation to orientedness usually improves denoising PSNR by a modest amount (.02–0.1 dB). In general, allowing adaptation to orientedness helps more for images with significant nonoriented content.

The adaptation to orientedness is mediated by the  $\delta$  binary hidden variable, whose prior distribution is determined for each subband by  $\beta$ , the probability of each patch arising from the oriented component. Values of the  $\beta$  parameter, estimated at noise level  $\sigma = 25$  and averaged over all subbands, are given in Table III. These estimated  $\beta$  values reflect how well the image data are described by the oriented versus the nonoriented model

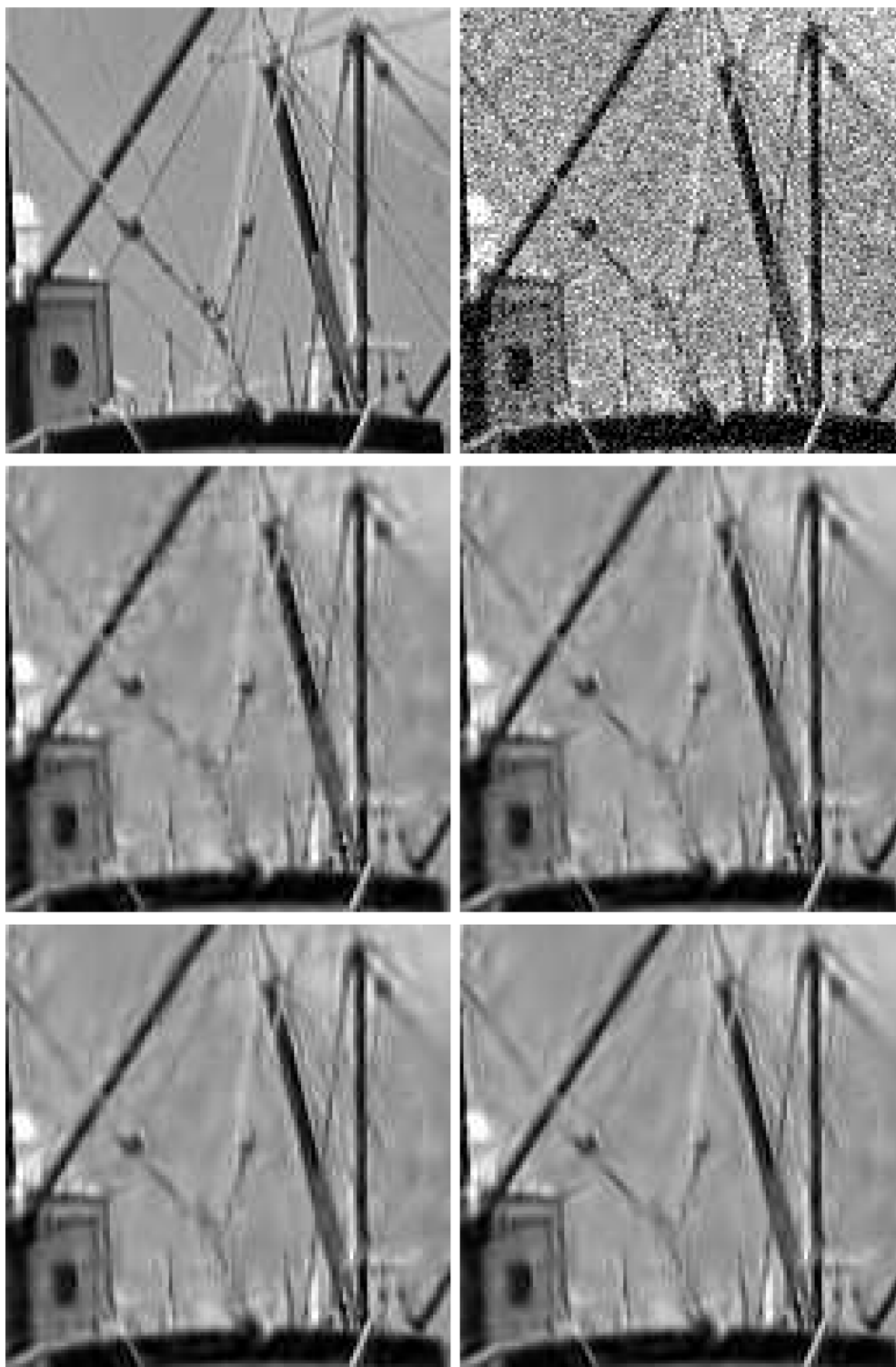


Fig. 6.  $128 \times 128$  detail from “Boats” image, for noise with  $\sigma = 25$ . Top: original, noisy (20.17). Middle: gsm2 (29.06 dB), oagsmnc2 (29.21 dB). Bottom: gsm8 (29.29 dB), oagsmnc8 (29.36 dB).

components, and thus intuitively provides a measure of the amount of oriented content in the image. For the test images shown, average  $\beta$  values were highest for the barbara image, which has significant oriented content such as the striped patterns on the shawl, and lowest for the boats image which contains significant nonoriented texture content. Additionally, the improvement in denoising performance of the OAGSM/NC

model over the standard GSM model is strongly correlated with the average  $\beta$ . Simple linear regression of this performance difference in dB versus average  $\beta$  for the five test images used, with  $\sigma = 25$  and two SP bands, gives a regression line with slope 1.24 ( $r^2 = 0.70$ ).

The OAGSM/NC denoising algorithm is significantly more computationally intensive than the GSM model. These costs can

TABLE III  
ESTIMATED VALUES OF  $\langle \beta \rangle$  AVERAGED OVER ALL TEN ORIENTED SUBBANDS,  
FOR 2-BAND PYRAMID WITH NOISE LEVEL  $\sigma = 25$ .  
STANDARD DEVIATIONS GIVEN IN PARENTHESIS

Lena	Barbara	Boats	House	Peppers
0.703 (0.19)	0.839 (0.07)	0.501 (0.18)	0.645 (0.23)	0.556 (0.24)

be separated into those due to parameter estimation and those due to MMSE estimation. Estimating the oriented covariance matrices by patch rotation is quite intensive. For the 2-band (8-band) algorithm we find the majority of the total computation time divided as follows: 40% (50%) for computing the oriented covariances, 25% (22%) for running the EM algorithm, and 28% (25%) for performing the MMSE estimation, with remaining time used for computing the forward and inverse SP transform and other overhead.

The MMSE estimator is computed as a numerical integral over the hidden variables  $z$ ,  $\theta$  and  $\delta$ . Integrating over  $z$  alone gives most of the computational cost for the GSM estimate. Integrating again over  $\theta$  and  $\delta$  would seem to imply this cost would increase by a factor equal to the product of the number of sample points for  $\theta$  and  $\delta$ . However, when  $\delta = 0$  the signal covariances do not depend on  $\theta$ , so integration over  $\theta$  is unnecessary. This implies that the cost of computing the MMSE estimate for the OAGSM/NC is  $N_\theta + 1$  times the cost of the original GSM model, where  $N_\theta$  is the number of sample values taken for the  $\theta$  hidden variable. On a 2.8 GHz AMD Opteron processor, using a nonoptimized Matlab implementation, denoising a  $512 \times 512$  pixel image takes approximately 10 minutes using the 2-band OAGSM/NC algorithm, and 60 minutes using the 8-band OAGSM/NC algorithm. Denoising a  $256 \times 256$  image took one quarter of these times.

#### D. Comparisons to Other Methods

We have compared the numerical results of our algorithm to four recent state-of-the-art methods, representing a variety of approaches to the image denoising problem. Amongst these are an algorithm based on a global statistical model, one based on sparse approximation methods, and two that are related to non-local averaging.

The Field of Experts model developed by Roth and Black [27] consists of a global Markov Random Field model, where the local clique potentials are given by products of simple “expert” distributions. These expert distributions are constructed as univariate functions of the response of a local patch to a particular filter. The structure of the model allows for each of these expert-generating filters to be trained from example data, yielding a flexible framework for learning image prior models. Denoising is then done with MAP estimation, by gradient ascent on the posterior probability density.

The KSVD method of Elad *et al.* [28] relies on the idea that natural images can be represented sparsely with an appropriate overcomplete set, or dictionary, of basis waveforms. Denoising is performed by using orthogonal matching pursuit to select an approximation of the noisy signal using a small number of dictionary elements. Intuitively, if the desired signal can be expressed sparsely, while the noise cannot, then the resulting approximation will preferentially recover the true signal. The

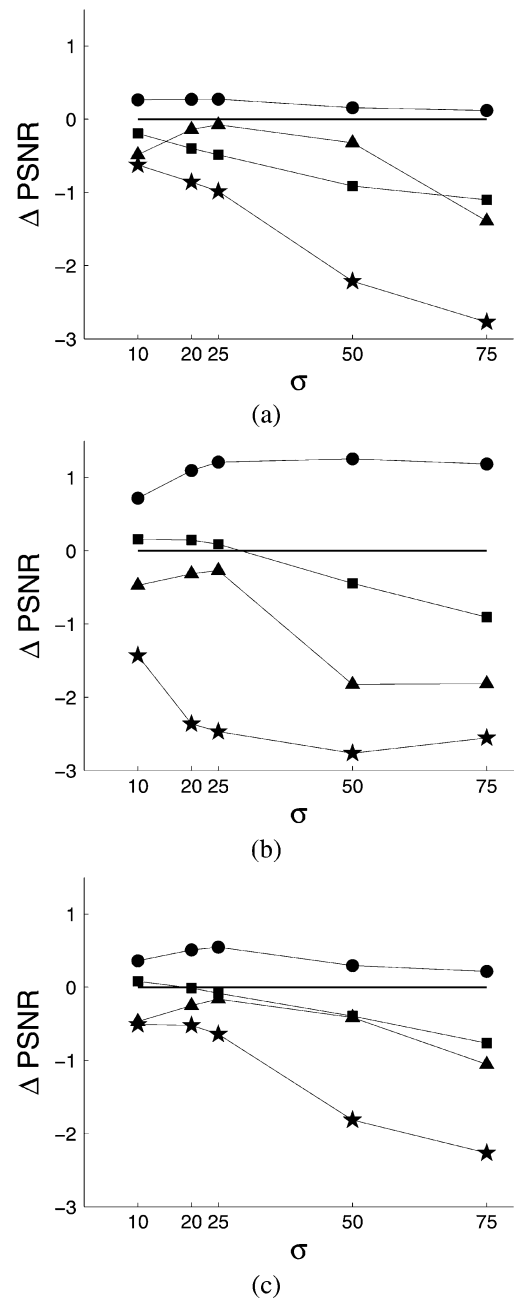


Fig. 7. Performance comparison of OAGSM/NC against other methods for (a) Lena, (b) Barbara, and (c) Boats images. Plotted points are differences in PSNR between various methods and the OAGSM/NC method. Circles (●) = 3D Collaborative Filtering, Dabov *et al.* [30], Squares (■) = KSVD, Elad and Aharon [28], Triangles (▲) = Kervrann and Boulanger [29], Stars (★) = Field of Experts, Roth and Black [27].

KSVD method gives a consistent framework for both learning an appropriate dictionary for image patches, and integrating the resulting local patch approximations into a global denoising method.

The work of Kervrann and Boulanger [29] constructs a denoising estimate where each pixel is computed as a weighted average over neighboring locations, where the weights are computed using a similarity measure of the surrounding patches. The success of this method relies on the property that similar, repeated patterns occur in areas nearby the current location to

be denoised. By averaging over such similar patches, noise is reduced while desired signal is retained. This paper also details a method for adaptively computing the size of the region over which neighboring patches are considered is at each location, allowing greater exploitation of repeated patterns where appropriate.

The collaborative filtering approach of Dabov *et al.* [30] also exploits this property, by explicitly grouping similar patches into 3-D “blocks” that are then jointly processed by wavelet shrinkage. Elements of the 3-D wavelet basis that correlate highly with features common across patches in the block will be preserved by the shrinkage operation, providing a powerful way of implicitly identifying and preserving structure across the patches. As the patches forming each such block may come from disparate spatial locations, this method can be understood as performing a type of nonlocal averaging. This method performs extremely well, producing, to our knowledge, the best PSNR results to date.

The differences in output PSNR between our OAGSM/NC algorithm and these four other denoising methods are shown graphically in Fig. 7. For the comparison with the KSVD method of Elad *et al.* [28], we used their publicly available code to compute denoised images from the same pseudorandom noise realizations used for the OAGSM/NC simulations. PSNR values for the other results were taken directly from the corresponding publications. We see that OAGSM/NC consistently outperforms all of the methods except for that of Dabov *et al.* [30].

The collaborative filtering approach of Dabov works by exploiting self similarity among image patches in different regions. While this approach yields very good denoising performance, the image structures that are restored through denoising are never explicitly described, but rather emerge implicitly through the grouping of similar patches. In contrast, the OAGSM/NC model attempts to much more explicitly describe the local image features. Rather than allowing descriptions of different image structures to emerge by partitioning into collections of similar patches, the patch rotation method for fitting the oriented covariances can be viewed as extracting an explicit description of a single oriented structure from patches at different orientations. As the OAGSM/NC is an explicit probabilistic model for image structure, its successful application for denoising helps provide insight into nature of the underlying image signal.

## V. CONCLUSIONS AND FUTURE WORK

We have introduced a novel OAGSM/NC image model that explicitly adapts to local image orientation. This model describes local patches of wavelet coefficients using a mixture of Gaussians, where the covariance matrices of the components are parametrized by hidden variables controlling the signal amplitude, orientation and orientedness. The model may be viewed as an extension of the Gaussian Scale Mixture model, which has a similar structure but with adaptation only to the local signal amplitude. We have developed methods for fitting the parameters of the OAGSM/NC model. A Bayes Least Squares optimal denoising estimator has been developed using

the OAGSM/NC model which shows noticeable improvement in both PSNR and visual quality compared to the GSM model, and compares favorably to several recent state-of-the-art denoising methods.

We believe that a number of other aspects of model implementation could be improved. One shortcoming for the current method of computing the oriented and nonoriented signal covariances is that we do not explicitly separate the contributions from the oriented and nonoriented processes. As the forward model describes each patch as a sample of either one process or the other, ideally the oriented covariances should be computed using only samples from the oriented process, and vice-versa. One possible way to account for this would be to estimate the signal covariances along with  $\beta$  inside of the EM procedure. Such a procedure is straightforward for a Gaussian mixture model, in which case the covariance matrices are given at each step as a weighted mixture of outer products of data points [26]. However, for the GSM and OAGSM densities direct application of EM to fit the covariances is more difficult, as the objective function then contains logarithms of sums where single covariance matrices appear multiple times with different scalar factors. For these densities, it is not the case that the maximum can be simply computed as a weighted average of outer products of the data points. It may nonetheless be possible to perform the M step for the full OAGSM/NC model by numerical gradient ascent.

As a preliminary step towards addressing this shortcoming, we have examined an ad-hoc scheme for computing the oriented and nonoriented covariances by weighting patches with the weights  $t_k^0$  and  $t_k^1$  from Section III-B. While this is not a true EM scheme for the reasons outlined above, it is intuitively appealing as it serves to softly partition the input patches into the oriented and nonoriented processes before computing the covariances. However, we found that this modification gave only very slight improvement in the denoising results (0.01–0.04 dB), at the expense of significant computational cost. Developing an efficient and principled method for computing the oriented and nonoriented covariances that better separates these two processes is an interesting question for further study.

Several other possible improvements remain. The current patch rotation method forms the oriented covariances using patches of actual image data that are not perfectly oriented. While this method gives good results, it would be interesting to consider more explicitly imposing the constraint on the oriented covariances that would arise from perfectly oriented, i.e., locally 1-D, signal patches. Another model aspect that could lead to improvements is the choice of neighborhood. Currently, co-localized patches from each oriented band are considered independently even though they represent information about the same localized features. Using neighborhoods that extend into adjacent orientation bands could thus improve performance. Another potentially worthwhile extension would be to incorporate additional hidden variables to capture signal characteristics other than energy, orientation and orientedness. For example, it should be possible to incorporate descriptions of local phase and local spatial frequency in an extension of the current OAGSM/NC model.

Finally, the model presented in this paper is a local model for image content, and does not take into account the interactions between patches. There is great potential for improvement in forming a consistent global model based on the local OAGSM/NC structure. The local Gaussian variables ( $u$ ) should be embedded in a Gaussian Markov Random field. This was recently done for the simple GSM local model [31], where it led to substantial improvements in denoising, albeit at significant computational cost. In addition, the hidden energy, orientation and orientedness variables are treated independently in the current model, despite the fact that they exhibit strong dependencies at adjacent locations, scales, and orientations. To account for such interactions, these hidden variables should also be embedded in random fields or trees, as has been done for GSMs [8], [32], [31].

## REFERENCES

- [1] W. K. Pratt, *Digital Image Processing*, 2nd ed. New York: Wiley Interscience, 1991.
- [2] S. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Pattern Anal. Mach. Intell.*, vol. 11, no. 7, pp. 674–693, 1989.
- [3] M. Antonini, T. Gaidon, P. Mathieu, and M. Barlaud, *Wavelets in Image Communication*. New York: Elsevier, 1994, ch. 3, pp. 65–188.
- [4] E. P. Simoncelli and E. H. Adelson, "Noise removal via bayesian wavelet coring," presented at the 3rd IEEE Int. Conf. Image Processing, 1996.
- [5] S. M. LoPresto, K. Ramchandran, and M. T. Orchard, "Wavelet image coding based on a new generalized Gaussian mixture model," presented at the Data Compression Conf, Snowbird, UT, Mar. 1997.
- [6] E. P. Simoncelli, "Statistical models for images: Compression, restoration and synthesis," in *Proc. 31st Asilomar Conf. Signals, Systems and Computers*, Pacific Grove, CA, Nov. 2–5, 1997, vol. 1, pp. 673–678.
- [7] M. S. Crouse, R. D. Nowak, and R. G. Baraniuk, "Wavelet-based statistical signal processing using hidden markov models," *IEEE Trans. Signal Process.*, vol. 46, no. 4, pp. 886–902, Apr. 1998.
- [8] M. Mihcak, I. Kozintsev, K. Ramchandran, and P. Moulin, "Low-complexity image denoising based on statistical modeling of wavelet coefficients," *IEEE Signal Process. Lett.*, vol. 6, no. 1, pp. 300–303, Jan. 1999.
- [9] M. J. Wainwright and E. P. Simoncelli, "Scale mixtures of Gaussians and the statistics of natural images," presented at the Advances in Neural Information Processing Systems (NIPS) 12, 2000.
- [10] M. Figueiredo and R. Nowak, "Wavelet-based image estimation: An empirical Bayes approach using Jeffrey's noninformative prior," *IEEE Trans. Image Process.*, vol. 10, no. 9, pp. 1322–1331, Sep. 2001.
- [11] A. Hyvärinen, J. Hurri, and J. Väyrynen, "Bubbles: A unifying framework for low-level statistical properties of natural image sequences," *J. Opt. Soc. America A*, vol. 20, no. 7, Jul. 2003.
- [12] Y. Karklin and M. S. Lewicki, "A hierarchical bayesian model for learning nonlinear statistical regularities in nonstationary natural signals," *Neural Comput.*, vol. 17, no. 2, pp. 397–423, Feb. 2005.
- [13] D. Andrews and C. Mallows, "Scale mixtures of normal distributions," *J. Roy. Statist. Soc. B*, vol. 36, pp. 99–102, 1974.
- [14] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli, "Image denoising using scale mixtures of Gaussians in the wavelet domain," *IEEE Trans. Image Process.*, vol. 12, no. 11, pp. 1338–1351, Nov. 2003.
- [15] J. A. Guerrero-Colón, L. Mancera, and J. Portilla, "Image restoration using space-variant Gaussian scale mixture in overcomplete pyramids," *IEEE Trans. Image Process.*, vol. 17, no. 1, pp. 27–41, Jan. 2008.
- [16] D. Hammond and E. Simoncelli, "Image denoising with an orientation-adaptive Gaussian scale mixture model," presented at the 13th IEEE Int. Conf. Image Processing, 2006.
- [17] E. P. Simoncelli, W. T. Freeman, E. H. Adelson, and D. J. Heeger, "Shiftable multi-scale transforms," *IEEE Trans. Inf. Theory*, vol. 38, no. 2, pp. 587–607, Mar. 1992, Special Issue on Wavelets.
- [18] J. M. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients," *IEEE Trans. Signal Processing*, vol. 41, no. 12, pp. 3445–62, Dec. 1993.
- [19] Z. Xiong, M. T. Orchard, and Y.-Q. Zhang, "A deblocking algorithm for JPEG compressed images using overcomplete wavelet representations," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 7, no. 2, pp. 433–437, Apr. 1997.
- [20] R. Buccigrossi and E. P. Simoncelli, "Image compression via joint statistical characterization in the wavelet domain," *IEEE Trans. Image Process.*, vol. 8, no. 12, pp. 1688–1701, Dec. 1999.
- [21] J. K. Romberg, H. Choi, and R. G. Baraniuk, "Bayesian tree-structured image modeling using wavelet-domain hidden markov models," *IEEE Trans. Image Process.*, vol. 10, no. 7, pp. 1056–1068, Jul. 2001.
- [22] J. Liu and P. Moulin, "Information-theoretic analysis of interscale and intrascale dependencies between image wavelet coefficients," *IEEE Trans. Image Process.*, vol. 10, no. 11, pp. 1647–1658, Nov. 2001.
- [23] D. L. Ruderman and W. Bialek, "Statistics of natural images: Scaling in the woods," *Phys. Rev. Lett.*, vol. 73, no. 6, pp. 814–817, 1994.
- [24] W. T. Freeman and E. H. Adelson, "The design and use of steerable filters," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 9, pp. 891–906, Sep. 1991.
- [25] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Statist. Soc. B*, vol. 39, pp. 1–38, 1977.
- [26] J. A. Bilmes, "A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden markov models," Tech. Rep. (TR-97-021), Dept. Elect. Eng. Comput. Sci., Univ. California, Berkeley, 1998.
- [27] S. Roth and M. Black, "Fields of experts: A framework for learning image priors," *Comput. Vis. Pattern Recognit.*, Jan. 2005.
- [28] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.*, vol. 15, no. 1, pp. 3736–3745, Jan. 2006.
- [29] C. Kervrann and J. Boulanger, "Optimal spatial adaptation for patch-based image denoising," *IEEE Trans. Image Process.*, vol. 15, no. 1, pp. 2866–2878, Jan. 2006.
- [30] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3D transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, no. 1, pp. 2080–2095, Jan. 2007.
- [31] S. Lyu and E. P. Simoncelli, "Modeling multiscale subbands of photographic images with fields of Gaussian scale mixtures," *IEEE Trans. Pattern Anal. Mach. Intell.*, 2008, to be published.
- [32] M. J. Wainwright, E. P. Simoncelli, and A. S. Willsky, "Random cascades on wavelet trees and their use in modeling and analyzing natural imagery," *Appl. Comput. Harmon. Anal.*, vol. 11, no. 1, pp. 89–123, Jul. 2001.
- [33] D. Taubman and A. Zakhor, "Orientation adaptive subband coding of images," *IEEE Trans. Image Process.*, vol. 3, no. 7, pp. 404–420, Jul. 1994.
- [34] X. Li, M. Orchard, and A. S. Willsky, "Edge directed prediction for lossless compression of natural images," *IEEE Trans. Image Process.*, vol. 10, no. 6, pp. 813–817, Jun. 2001.



**David K. Hammond** was born in Loma Linda, CA. He received the B.S. degree with honors in mathematics and chemistry from Caltech, Pasadena, in 1999, and the Ph.D. degree in mathematics from the Courant Institute of Mathematical Sciences, New York University, in 2007.

He served as a Peace Corps volunteer teaching secondary mathematics in Malawi from 1999–2001. In September 2007, he joined the Ecole Polytechnique Federale de Lausanne, France, where he is currently a Postdoctoral Researcher. His current research interests focus on image processing and statistical signal models, including applications to astronomical data analysis.



**Eero P. Simoncelli** (S'92–M'93–SM'04) received the B.S. degree (summa cum laude) in physics from Harvard University, Cambridge, MA, in 1984. He studied applied mathematics at the University of Cambridge, Cambridge, U.K., for a year and a half, and then received the M.S. and Ph.D. degrees in electrical engineering from the Massachusetts Institute of Technology, Cambridge, MA, in 1988 and 1993, respectively.

He was an Assistant Professor in the Computer and Information Science department, University of Pennsylvania, Philadelphia, from 1993 until 1996. He joined New York University in September of 1996, where he is currently a Professor of neural science and mathematics. In August 2000, he became an Investigator of the Howard Hughes Medical Institute, under their new program in computational biology. His research interests span a wide range of topics in the representation and analysis of visual images, in both machine and biological systems.