

# IMAGE RESTORATION USING GAUSSIAN SCALE MIXTURES IN THE WAVELET DOMAIN

Javier Portilla \*

Visual Information Processing Group  
Dept. of Comp. Science and Artif. Intell.  
Universidad de Granada  
javier@decsai.ugr.es

Eero Simoncelli †

Center for Neural Science, and  
Courant Inst. of Mathematical Sciences  
New York University, NY 10003  
eero.simoncelli@nyu.edu

*We describe a statistical model for images decomposed in an overcomplete wavelet pyramid. Each neighborhood of pyramid coefficients is modeled as the product of a Gaussian vector of known covariance, and an independent hidden positive scalar random variable. We propose an efficient Bayesian estimator for the pyramid coefficients of an image degraded by linear distortion (e.g., blur) and additive Gaussian noise. We demonstrate the quality of our results in simulations over a wide range of blur and noise levels.*

## 1. INTRODUCTION

Natural images have distinct features that allow the human visual system to detect the presence of distortion, and to extract remaining information from the observation. Image restoration aims to construct an approximation sharing the relevant features still present in the corrupted image, but with the artifacts suppressed. In order to distinguish the artifacts from the signal, a good image model is essential. In this work we use a Bayesian framework for image restoration, basing on a prior statistical model for natural images. We assume Gaussian noise of known covariance has been added after having convolved the image with a known blurring kernel. Such model for the distortion provides a reasonable approximation to real-world corruption sources, such as de-focus or photon and electron noise in cameras. In previous works, we have described a model for wavelet coefficients using scale mixtures of Gaussians [1, 2]. Related models have been developed by several groups [3, 4, 5, 6]. We have applied this model to estimate images in the presence of independent additive Gaussian noise of known covariance, following a suboptimal empirical Bayes strategy [2, 7]. Recently, we developed a direct Bayesian Least Square denoising procedure [8]. In this paper, we extend this approach to include the full restoration problem.

\*JP is supported by a "Ramón y Cajal" grant (Science and Technology Spanish Ministry)

†EPS is supported by NSF CAREER grant MIP-9796040.

## 2. IMAGE MODELING

Modeling the statistics of natural images is a difficult task, because of the high dimensionality of the signal and the higher-order statistical dependencies that are prevalent. Simplifying assumptions, such as homogeneity and locality are usually applied. In the past decade it has become standard to begin by decomposing the image with a set of multi-scale band-pass oriented filters. This kind of representation is adapted to the approximate scale invariance of natural images, and it has been shown to decouple some high-order statistics. Another technique for simplifying the description of high-order statistics is to use a low-order local model (e.g., Gaussian) whose parameters (e.g., variance) are governed by a local hidden random variable. In this work we have combined these ideas into a Bayesian framework.

### 2.1. Image representation

As a preprocessing stage for image modeling, we use a fixed multi-scale multi-orientation linear decomposition. It is now generally agreed that the use of overcomplete representations is advantageous for restoration, in order to avoid aliasing artifacts that plague critically-sampled representations such as orthogonal wavelets [9]. A widely followed approach is to use orthogonal or biorthogonal basis functions, but without decimating the subbands [e.g., 10]. However, after the critical sampling constraint has been removed, significant improvement comes from using representations with higher redundancy and orientation selectivity [9, 11, 12]. For the current paper, we have used a particular variant of an overcomplete tight frame representation known as a "steerable pyramid" [9] (see [8] for details). In this paper we have used 5 scales and 4 orientations for the decomposition, for a total number of 25 subbands. This number includes  $4 \times 5$  band-pass subbands plus 4 oriented high-pass subbands and a non-oriented low-pass residual.

## 2.2. Gaussian scale mixtures in the wavelet domain

For each coefficient in the pyramid representation, we consider a *neighborhood* of coefficients, referring to the center coefficient as the *reference coefficient* of the neighborhood. The neighborhood may include coefficients from other subbands (i.e., corresponding to basis functions at nearby scales and orientations), as well as from the same subband. For this work we have used a  $3 \times 3$  neighborhood around the reference coefficient, plus the *parent* coefficient (same orientation and position, next coarser scale), whenever it exists. Due to sampling differences at different scales, the coarser subband must be resampled at double rate in both dimensions for obtaining the parent of each coefficient of the subband.

We use a Gaussian scale mixture (GSM) to model the coefficients within each local neighborhood with reference coefficients belonging to every given subband. A random vector  $\mathbf{x}$  is a Gaussian scale mixture [13] if it can be expressed as the product of a zero-mean Gaussian vector  $\mathbf{u}$  and an independent positive scalar random variable  $\sqrt{z}$ :

$$\mathbf{x} = \sqrt{z}\mathbf{u}. \quad (1)$$

The variable  $z$  is the multiplier. Vector  $\mathbf{x}$  is an infinite mixture of Gaussian vectors, whose density is determined by the covariance matrix  $\mathbf{C}_{\mathbf{uu}}$  of  $\mathbf{u}$  and the mixing density  $p_z(z)$ :

$$\begin{aligned} p_{\mathbf{x}}(\mathbf{x}) &= \int p(\mathbf{x}|z)p_z(z) dz \\ &= \int \frac{\exp(-\mathbf{x}^T(z\mathbf{C}_{\mathbf{uu}})^{-1}\mathbf{x}/2)}{(2\pi)^{N/2}|z\mathbf{C}_{\mathbf{uu}}|^{1/2}} p_z(z) dz, \end{aligned} \quad (2)$$

where  $N$  is the dimensionality of  $\mathbf{x}$  and  $\mathbf{u}$  (the size of the neighborhood). Without loss of generality one can assume  $\mathbb{E}\{z\} = 1$ , which implies  $\mathbf{C}_{\mathbf{uu}} = \mathbf{C}_{\mathbf{xx}}$ .

GSM densities are symmetric and zero-mean, and they have leptokurtotic marginal densities (i.e., heavier tails than a Gaussian). Another key property of the GSM model is that the density of  $\mathbf{x}$  is Gaussian when conditioned on  $z$ . Also, the normalized vector  $\mathbf{x}/\sqrt{z}$  is Gaussian. They also present interesting joint statistics: the variance of a vector element conditioned on a neighbor scales roughly linearly with the square of the neighbor's value. These marginal and joint statistics of GSM distributions are qualitatively similar to those of neighbor coefficients responding to natural images in multi-scale and multi-orientation representation [14, 1].

## 2.3. Prior density for multiplier

Our model requires a specification of the density of the multiplier,  $z$ . An alternative to estimate adaptive priors for the hidden multiplier (as in [e.g., 5, 7]) is to use a *noninformative prior* [15], which does not require the fitting of any parameters to the observation. We have applied *Jeffrey's*

*prior*, which for a Gaussian variable with a scale parameter  $\sqrt{z}$  yields [16]:

$$p_z(z) \propto \frac{1}{z}. \quad (3)$$

To avoid computational problems at the estimation stage, we have set the prior to zero in the interval  $[0, z_{\min})$ , where  $z_{\min}$  is a small positive number.<sup>1</sup> This solution is simple and efficient to implement. Surprisingly, it also performs better than maximum likelihood fitted non-parametric priors for the neighborhoods of each subband.<sup>2</sup>

## 3. RESTORATION

We model the observed image as:

$$y(n, m) = x(n, m) * h(n, m) + w(n, m), \quad (4)$$

where  $*$  denotes convolution,  $x$  is the original image,  $h$  is a known linear kernel and  $w$  is Gaussian noise of known power spectral density (PSD)  $P_w(u, v)$ .

### 3.1. GSM with blur and noise

We follow the usual procedure for image restoration in the wavelet domain: (1) decompose the image  $y$  into pyramid subbands; (2) restore each subband, except for the low-pass residual; and (3) invert the pyramid transform. Translating (4) into our local GSM model, a vector  $\mathbf{y}$  corresponding to a neighborhood of  $N$  observed coefficients around a reference coefficient for each subband of the pyramid representation can be expressed as:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w} = \sqrt{z}\mathbf{H}\mathbf{u} + \mathbf{w}, \quad (5)$$

where  $\mathbf{u}$  and  $\mathbf{w}$  are zero-mean Gaussian vectors, and  $\mathbf{H}$  is an  $N \times M$  matrix expressing the convolution of  $N$  coefficients with a kernel  $h$  of size  $N_h$  ( $N_h \gg N$ , typically).<sup>3</sup> The density of the observed neighborhood vector conditioned on  $z$  is zero-mean Gaussian, with covariance  $\mathbf{C}_{\mathbf{y}|z} = z\mathbf{C}_{\mathbf{u}'\mathbf{u}'} + \mathbf{C}_{\mathbf{ww}}$ :

$$p(\mathbf{y}|z) = \frac{\exp(-\mathbf{y}^T(z\mathbf{C}_{\mathbf{u}'\mathbf{u}'} + \mathbf{C}_{\mathbf{ww}})^{-1}\mathbf{y}/2)}{\sqrt{(2\pi)^N |z\mathbf{C}_{\mathbf{u}'\mathbf{u}'} + \mathbf{C}_{\mathbf{ww}}|}}, \quad (6)$$

where  $\mathbf{C}_{\mathbf{u}'\mathbf{u}'} = \mathbf{H}\mathbf{C}_{\mathbf{uu}}\mathbf{H}^T$  is the  $N \times N$  covariance matrix of  $\mathbf{u}' = \mathbf{H}\mathbf{u}$ . As the noise PSD  $P_w(u, v)$  is assumed known,  $\mathbf{C}_{\mathbf{ww}}$  is easily estimated by computing, for each subband, the sample covariance for the pyramid coefficients

<sup>1</sup>We have chosen  $z_{\min} = 10^{-9}$  in our implementation. Results remain almost unaffected with  $z_{\min} \in [10^{-17}, 10^{-5}]$ .

<sup>2</sup>Note that, when estimating parameters under an overcomplete representation, the set of least squares optimal solutions for the subbands is not least square optimal for the whole image.

<sup>3</sup>The kernel  $h$  must be cropped in the frequency domain according to the spatial resolution of the pyramid subband.

of the inverse Fourier transform of  $\sqrt{P_w(u, v)}$  (a delta for the case of white noise). Given  $\mathbf{C}_{\mathbf{w}\mathbf{w}}$  and taking the expectation of  $\mathbf{C}_{\mathbf{y}|z}$  over  $z$ , the filtered signal covariance  $\mathbf{C}_{\mathbf{u}'\mathbf{u}'}$  can be computed from the observation covariance matrix  $\mathbf{C}_{\mathbf{y}\mathbf{y}}$ :  $\mathbf{C}_{\mathbf{y}\mathbf{y}} = \mathbb{E}\{z\}\mathbf{C}_{\mathbf{u}'\mathbf{u}'} + \mathbf{C}_{\mathbf{w}\mathbf{w}}$ . Setting  $\mathbb{E}\{z\}$  to 1, results in  $\mathbf{C}_{\mathbf{u}'\mathbf{u}'} = \mathbf{C}_{\mathbf{x}'\mathbf{x}'} = \mathbf{C}_{\mathbf{y}\mathbf{y}} - \mathbf{C}_{\mathbf{w}\mathbf{w}}$ . To ensure positive-definiteness, small negative eigenvalues are set to zero.

### 3.2. Bayes least squares estimator

For each neighborhood, we estimate  $x_c$ , the reference coefficient, from  $\mathbf{y}$ , the set of noisy coefficients. The Bayes least squares (BLS) estimate is:

$$\begin{aligned} \mathbb{E}\{x_c|\mathbf{y}\} &= \int x_c p(x_c|\mathbf{y}) dx_c \\ &= \int \int_0^\infty x_c p(x_c|\mathbf{y}, z) p(z|\mathbf{y}) dz dx_c \\ &= \int_0^\infty p(z|\mathbf{y}) \mathbb{E}\{x_c|\mathbf{y}, z\} dz. \end{aligned} \quad (7)$$

Thus, the solution is an average of the least squares estimate of  $x_c$  when conditioned on  $z$  (local Wiener solution), weighted by the posterior density of the multiplier,  $p(z|\mathbf{y})$ . This integral can be computed numerically for each neighborhood of coefficients with a few (10 to 15) uniform samples in  $\log z$ . We now describe each of these components.

### 3.3. Local Wiener estimate

The local linear estimate for the full neighborhood is the Wiener solution:

$$\mathbb{E}\{\mathbf{x}|\mathbf{y}, z\} = z\mathbf{C}_{\mathbf{x}\mathbf{x}'} (z\mathbf{C}_{\mathbf{x}'\mathbf{x}'} + \mathbf{C}_{\mathbf{w}\mathbf{w}})^{-1} \mathbf{y}, \quad (8)$$

where  $\mathbf{C}_{\mathbf{x}\mathbf{x}'} = \mathbf{C}_{\mathbf{x}\mathbf{x}}\mathbf{H}^T$  is the  $M \times N$  cross-covariance matrix of  $\mathbf{x}$  and  $\mathbf{x}' = \mathbf{H}\mathbf{x}$ , the coefficients from the original image and those from its blurry version. We explain below a method for estimating this matrix.

We can simplify the dependence of this expression on  $z$  by diagonalizing the matrix  $z\mathbf{C}_{\mathbf{x}'\mathbf{x}'} + \mathbf{C}_{\mathbf{w}\mathbf{w}}$ . Specifically, let  $\mathbf{S}$  be the symmetric square root of the positive definite matrix  $\mathbf{C}_{\mathbf{w}\mathbf{w}}$  (i.e.,  $\mathbf{C}_{\mathbf{w}\mathbf{w}} = \mathbf{S}\mathbf{S}^T$ ), and  $\{\mathbf{Q}, \mathbf{\Lambda}\}$  the eigenvector/eigenvalue expansion of the matrix  $\mathbf{S}^{-1}\mathbf{C}_{\mathbf{x}'\mathbf{x}'}\mathbf{S}^{-T}$ . Then:

$$\begin{aligned} z\mathbf{C}_{\mathbf{x}'\mathbf{x}'} + \mathbf{C}_{\mathbf{w}\mathbf{w}} &= z\mathbf{C}_{\mathbf{x}'\mathbf{x}'} + \mathbf{S}\mathbf{S}^T \\ &= \mathbf{S} (z\mathbf{S}^{-1}\mathbf{C}_{\mathbf{x}'\mathbf{x}'}\mathbf{S}^{-T} + \mathbf{I}) \mathbf{S}^T \\ &= \mathbf{S}\mathbf{Q} (z\mathbf{\Lambda} + \mathbf{I}) \mathbf{Q}^T \mathbf{S}^T. \end{aligned} \quad (9)$$

This diagonalization does not depend on  $z$ , and thus only needs to be computed once for each subband. We can now simplify (8):

$$\begin{aligned} \mathbb{E}\{\mathbf{x}|\mathbf{y}, z\} &= z\mathbf{C}_{\mathbf{x}\mathbf{x}'}\mathbf{S}^{-T}\mathbf{Q}(z\mathbf{\Lambda} + \mathbf{I})^{-1}\mathbf{Q}^T\mathbf{S}^{-1}\mathbf{y} \\ &= z\mathbf{M}(z\mathbf{\Lambda} + \mathbf{I})^{-1}\mathbf{v}, \end{aligned} \quad (10)$$

where  $\mathbf{M} = \mathbf{C}_{\mathbf{x}\mathbf{x}'}\mathbf{S}^{-T}\mathbf{Q}$ , and  $\mathbf{v} = \mathbf{Q}^T\mathbf{S}^{-1}\mathbf{y}$ . Finally, we restrict the estimate to the reference coefficient:

$$\mathbb{E}\{x_c|\mathbf{y}, z\} = \sum_{n=1}^N \frac{zm_{cn}v_n}{z\lambda_n + 1}, \quad (11)$$

where  $m_{ij}$  represents an element ( $i$ -th row,  $j$ -th column) of the matrix  $\mathbf{M}$ ,  $\lambda_n$  are the diagonal elements of  $\mathbf{\Lambda}$ ,  $v_n$  the elements of  $\mathbf{v}$ , and  $c$  is the index of the reference coefficient within the neighborhood vector.

#### 3.3.1. Estimation of the cross-covariance matrix

We use the relation between the power spectrum of the observed image and that of the original image,  $P_y(u, v) = P_x(u, v)|H(u, v)|^2 + P_w(u, v)$ , to first obtain an approximation of the power spectrum of the filtered image:

$$P_{x'}(u, v) \simeq [|Y(u, v)|^2 * G(u, v) - P_w(u, v)]_+, \quad (12)$$

where  $G(u, v)$  is a Gaussian convolving window for removing sampling fluctuations. Then, we use:

$$P_x(u, v) \simeq P_{x'}(u, v) / \max(|H(u, v)|^2, G_{max}^{-1}), \quad (13)$$

where  $G_{max}$  is the maximal allowed gain for the inverse filter. For the results shown in this paper (white noise case) we chose  $G_{max} = \min(2400/\sigma_0^2, 180)$  with  $\sigma_0$  the standard deviation of  $w$ .<sup>4</sup> Now  $\mathbf{C}_{\mathbf{x}\mathbf{x}'}$  can be estimated as the sample cross-covariance of the coefficients of the inverse Fourier transforms of  $\sqrt{P_x(u, v)}$  and  $\sqrt{P_{x'}(u, v)}$ . The resulting estimate  $\widehat{\mathbf{C}_{\mathbf{x}\mathbf{x}'}}$  can be significantly improved by modeling its bias using an  $N \times N$  linear transform:  $\mathbb{E}\{\widehat{\mathbf{C}_{\mathbf{x}\mathbf{x}'}}\} \simeq \mathbf{C}_{\mathbf{x}\mathbf{x}'}\mathbf{B}$ . In order to estimate  $\mathbf{B}$ , we assume that we would incur the same proportional bias when estimating  $\mathbf{C}_{\mathbf{x}'\mathbf{x}'}$  from  $P_{x'}(u, v)$ :  $\mathbb{E}\{\widehat{\mathbf{C}_{\mathbf{x}'\mathbf{x}'}}\} \simeq \mathbf{C}_{\mathbf{x}'\mathbf{x}'}\mathbf{B} = (\mathbf{C}_{\mathbf{y}\mathbf{y}} - \mathbf{C}_{\mathbf{w}\mathbf{w}})\mathbf{B}$ . Solving this for  $\mathbf{B}^{-1}$  yields the following improved estimator:

$$\widehat{\widehat{\mathbf{C}_{\mathbf{x}\mathbf{x}'}}} = \widehat{\mathbf{C}_{\mathbf{x}\mathbf{x}'}} \left[ (\mathbf{C}_{\mathbf{y}\mathbf{y}} - \mathbf{C}_{\mathbf{w}\mathbf{w}})\widehat{\mathbf{C}_{\mathbf{x}'\mathbf{x}'}}^{-1} \right] \quad (14)$$

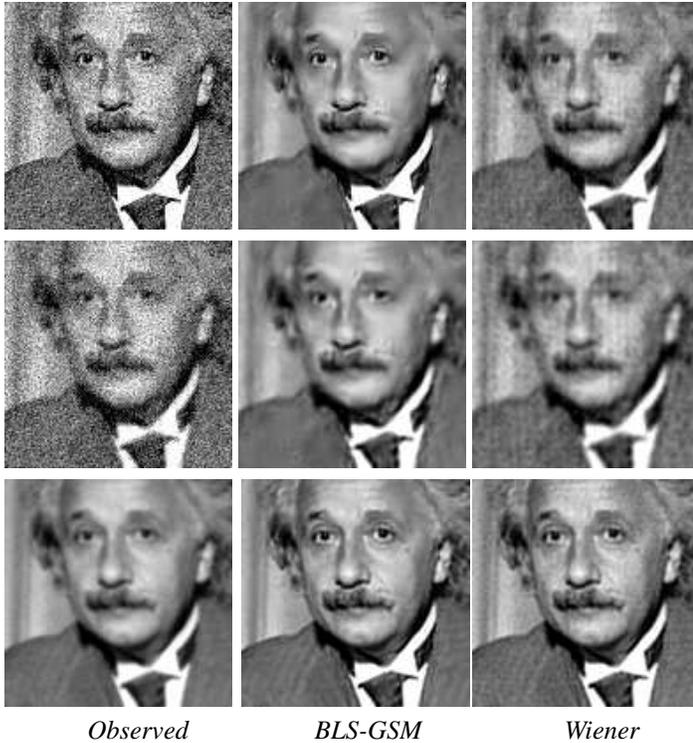
Note that particularizing (14) for the reference  $c$ -th row of  $\mathbf{C}_{\mathbf{x}\mathbf{x}'}$  ( $N$  values) suffices for solving (11).

### 3.4. Posterior distribution of the multiplier

We use Bayes rule to compute  $p(z|\mathbf{y})$  from  $p(\mathbf{y}|z)$  and  $p(z)$ , normalizing numerically the integral of their product. Using the relationship in (9) and the definition of  $\mathbf{v}$ , the conditional density  $p(\mathbf{y}|z)$  defined in (6) simplifies to:

$$p(\mathbf{y}|z) = \frac{\exp(-\frac{1}{2} \sum_{n=1}^N \frac{v_n^2}{z\lambda_n + 1})}{\sqrt{(2\pi)^N |\mathbf{C}_{\mathbf{w}\mathbf{w}}| \prod_{n=1}^N (z\lambda_n + 1)}}. \quad (15)$$

<sup>4</sup>Other more accurate estimates could be used instead.



**Fig. 1.** **Up:** High noise ( $\sigma = 20$ ) and low blur ( $\sigma_b = 0.2$ ); PSNR values are (dB): 22.1, 29.9, 28.3; **Middle:** High noise ( $\sigma = 20$ ) and high blur ( $\sigma_b = 2.0$ ), (21.5, 27.9, 27.0); **Bottom:** Low noise ( $\sigma = 2$ ) and high blur ( $\sigma_b = 2.0$ ), (29.7, 32.5, 31.8);

#### 4. SIMULATIONS

We have applied our restoration algorithm to the *Einstein* image. For  $h$  we used a unity volume circular low-pass Gaussian filter with standard deviation  $\sigma_b$ . We added zero-mean white Gaussian noise of variance  $\sigma_0^2$ . We compare to the Wiener linear restoration method, (using (12) and (13)). Fig. 1 shows the results with 2 levels of noise and blur.

#### 5. REFERENCES

- [1] M J Wainwright and E P Simoncelli, "Scale mixtures of Gaussians and the statistics of natural images," in *Adv. Neural Information Processing Systems*, S. A. Solla, T. K. Leen, and K.-R. Müller, Eds., Cambridge, MA, May 2000, vol. 12, pp. 855–861, MIT Press,
- [2] M Wainwright, E Simoncelli, and A Willsky, "Random cascades on wavelet trees and their use in modeling and analyzing natural imagery," *Applied and Computational Harmonic Analysis*, 2000,
- [3] S M LoPresto, K Ramchandran, and M T Orchard, "Wavelet image coding based on a new generalized gaussian mixture model," in *Data Compression Conf.*, Snowbird, Utah, 1997.
- [4] M S Crouse, R D Nowak, and R G Baraniuk, "Wavelet-based statistical signal processing using hidden Markov models," *IEEE Trans. Signal Proc.*, vol. 46, pp. 886–902, April 1998.
- [5] M K Mihçak, I Kozintsev, K Ramchandran, and P Moulin, "Low-complexity image denoising based on statistical modeling of wavelet coefficients," *IEEE Trans. Sig. Proc.*, vol. 6, no. 12, pp. 300–303, December 1999.
- [6] C Spence and L Parra, "Hierarchical image probability (HIP) model," in *Adv. Neural Information Processing Systems*, S. A. Solla, T. K. Leen, and K.-R. Müller, Eds., Cambridge, MA, May 2000, vol. 12, MIT Press.
- [7] J Portilla, V Strela, M Wainwright, and E Simoncelli, "Adaptive Wiener denoising using a Gaussian scale mixture model in the wavelet domain," in *Proc 8th IEEE Int'l Conf on Image Proc.*, Thessaloniki, Greece, 2001, pp. 37–40
- [8] J Portilla, V Strela, M Wainwright, and E P Simoncelli, "Image denoising using scale mixtures of Gaussians in the wavelet domain," *IEEE Trans. Image Proc.*, In Press. 2003.
- [9] E P Simoncelli, W T Freeman, E H Adelson, and D J Heeger, "Shiftable multi-scale transforms," *IEEE Trans Information Theory*, vol. 38, no. 2, pp. 587–607, March 1992,
- [10] R R Coifman and D L Donoho, "Translation-invariant denoising," in *Wavelets and statistics*, A Antoniadis and G Oppenheim, Springer-Verlag lecture notes, San Diego, 1995.
- [11] E P Simoncelli, "Bayesian denoising of visual images in the wavelet domain," in *Bayesian Inference in Wavelet Based Models*, P Müller and B Vidakovic, Eds., chapter 18, pp. 291–308. Springer-Verlag, New York, Spring 1999,
- [12] J Starck, E J Candes, and D L Donoho, "The curvelet transform for image denoising," *IEEE Trans. Image Proc.*, vol. 11, no. 6, pp. 670–684, June 2002.
- [13] D Andrews and C Mallows, "Scale mixtures of normal distributions," *J. Royal Stat. Soc.*, vol. 36, pp. 99–, 1974.
- [14] E P Simoncelli, "Statistical models for images: Compression, restoration and synthesis," in *31st Asilomar Conf on Signals, Systems and Computers*, Pacific Grove, CA, November 1997, pp. 673–678, IEEE Computer Society,
- [15] M Figueiredo and R Nowak, "Bayesian wavelet-based image estimation using non-informative priors," in *SPIE Conf. on Mathematical Modeling, Bayesian Estimation, and Inverse Problems*, Denver CO, Jul 1999.
- [16] J O Berger, "Prior information and subjective probability," in *Statistical Decision Theory and Bayesian Analysis*, Springer Series in Statistics. Springer-Verlag, 1990.