Published in: Proc. SPIE 45th Annual Meeting
San Diego, California. July 30 - August 4, 2000.

Image Denoising Using a Local Gaussian Scale Mixture Model in the Wavelet Domain

Vasily Strela[a], Javier Portilla[b], and Eero Simoncelli[b]

[a]Dept. of Mathematics and Computer Science,
Drexel University, Philadelphia, PA 19104, USA

[b]Ctr. for Neural Science, and Courant Inst. of Mathematical Sciences,
New York University, New York, NY 10003, USA

ABSTRACT

The statistics of photographic images, when decomposed in a multiscale wavelet basis, exhibit striking non-Gaussian behaviors. The joint densities of clusters of wavelet coefficients (corresponding to basis functions at nearby spatial positions, orientations and scales) are well-described as a Gaussian scale mixture: a jointly Gaussian vector multiplied by a hidden scaling variable. We develop a maximum likelihood solution for estimating the hidden variable from an observation of the cluster of coefficients contaminated by additive Gaussian noise. The estimated hidden variable is then used to estimate the original noise-free coefficients. We demonstrate the power of this model through numerical simulations of image denoising.

Keywords: natural image statistics, wavelet, multiresolution, Gaussian scale mixture, adaptive Wiener filtering, denoising

1. INTRODUCTION

Most applications in image processing can benefit from a good statistical model (i.e., a prior) for the relevant set of images. In recent years, multiscale wavelet decompositions have led to the development of a new generation of statistical image models. Wavelet coefficients of images show remarkably regular but non-Gaussian properties, both in their marginal[1,2] and their joint[3,4] statistics. In particular, the *magnitudes* of nearby wavelet coefficients of photographic images are often strongly correlated, even when the raw coefficients are decorrelated.[4] Wavelet marginal models are the basis for shrinkage or coring methods of denoising.[5–7] A number of recent approaches to compression and denoising take advantage of joint statistical relationships by adaptively estimating the variance of a coefficient from a local neighborhood consisting of coefficients within a subband,[8–11] or including coefficients from subbands adjacent in scale and/or orientation.[12–14] This kind of local variance estimation was originally developed in the context of adaptive Wiener denoising in the pixel domain.[15,16]

Wainwright and Simoncelli have proposed a model to capture both the marginal and joint statistical properties of local neighborhoods of coefficients, based on the semi-parametric class of random variable known as a *Gaussian scale mixture*[17] (GSM). In the GSM model, wavelet coefficients in a neighborhood correspond to a product of a Gaussian random vector with a hidden multiplier variable. Closely related models have been independently proposed by LoPresto *et. al.*[8] and Crouse *et. al.*.[18] Wainwright *et. al.* have extended the local GSM model to a global description using a coarse-to-fine cascade on the wavelet tree.[19,20]

In this paper, we forego the complexity of the global prior model, and develop maximum likelihood estimators for the parameters of a local GSM model. We demonstrate the use of these estimators in the problem of denoising an image that has been contaminated with additive Gaussian noise of known covariance.

Author e-mail addresses: vstrela@mcs.drexel.edu, javier@cns.nyu.edu, eero.simoncelli@nyu.edu

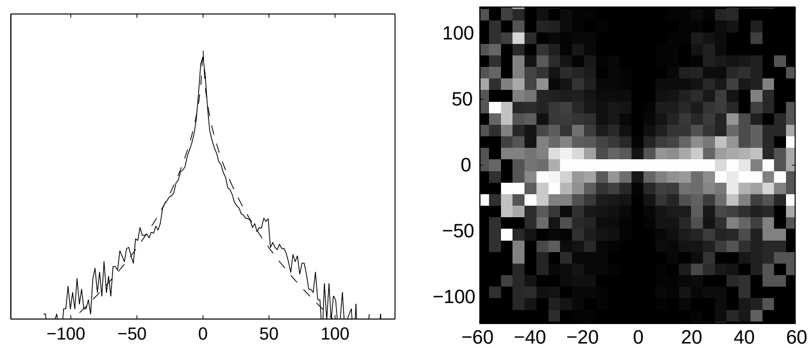


Figure 1. Example wavelet coefficient histograms of the "Boats" image. Left: Log marginal histogram of a single vertical subband. Dashed line is a generalized Gaussian density, with parameters chosen to minimize the relative entropy with the histogram.[6,12] Right: conditional joint histogram of two vertical coefficients at adjacent scales. Pixel brightness corresponds to the frequency of occurrence of the corresponding pair of coefficient values, except that each column has been independently rescaled to fill the full range of intensities. Note the dependence between the two coefficients: the variance of the ordinate variable grows with the magnitude of the abscissa variable.

2. LOCAL GAUSSIAN SCALE MIXTURE MODEL

Wavelet coefficients of photographic images exhibit strongly non-Gaussian heavy-tailed marginal distributions,[1,2] and these have been modeled using Laplacian or generalized Laplacian densities.[6,7] In addition, the joint densities of wavelet coefficients exhibit striking non-Gaussian dependencies, in which the variance of each coefficient is proportional to the squared magnitudes of coefficients at adjacent spatial positions, orientations, and scales.[4,21] Examples of both of these non-Gaussian behaviors are shown in figure 1.

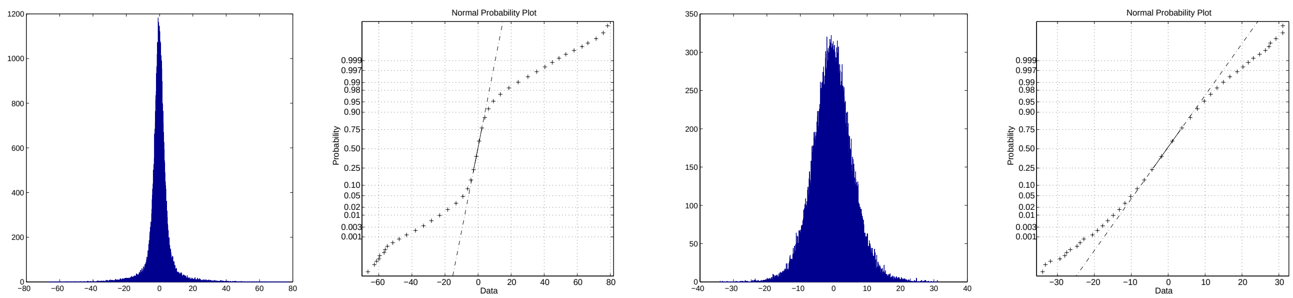
Wainwright and Simoncelli proposed the use of Gaussian scale mixtures as a model of these joint wavelet coefficient distributions.[17] A vector X is a GSM if it may be decomposed into a product of two random variables, $X = \sqrt{z}U$, where z is a positive scalar random variable and U is a zero-mean Gaussian random vector with covariance matrix C_u, independent of z.[22] The probability density of a GSM variable is :

$$\mathcal{P}_x(X) = \int \frac{1}{(2\pi)^{N/2}\ |zC_u|^{1/2}}\ \exp\left(\frac{-X^T C_u^{-1} X}{2z}\right)\ \mathcal{P}_z(z)\ dz,$$

where N is the dimensionality of vectors U and X. Note that if the components of U are decorrelated, then the components of X will also be decorrelated. The GSM class includes a number of well-known heavy tailed distributions such as generalized Gaussians, the α-stable family, and the Student t-variables.[17] The key advantage of these models is that the density of X is jointly Gaussian when conditioned on z. Specifically, the *normalized* quantity $X/\sqrt{z}$ is Gaussian distributed, with covariance C_u.

We assume a GSM model for local clusters of wavelet coefficients. Specifically, we assume that a vector containing wavelet coefficients corresponding to basis functions at nearby positions, orientations and scales is distributed as a GSM. For this paper, we have used clusters containing only spatial neighbors within the same subband. For simplicity, we will assume that the multipliers associated with different clusters are independent, even though the clusters are overlapping (i.e., contain common coefficients).

In order to test the ability of the GSM model to account for the statistics of natural images, we estimate multiplier z locally for each neighborhood in a wavelet subband. Let X be a vector containing a cluster of



a) raw (unnormalized) b) normalized by $\sqrt{\hat{z}(X)}$

Figure 2. Histogram and normal probability plots of vertical detail subband of Lenna image before and after normalization of each coefficient by the estimate of hidden multiplier $\sqrt{\hat{z}}$.

wavelet coefficients, with x the coefficient at the center of the cluster. The maximum likelihood estimator for the multiplier is[17,10]:

$$\hat{z}(X) = \frac{X^T C_u^{-1} X}{N}.$$

Given the maximum likelihood estimate $\hat{z}(X)$, we can form the *normalized coefficient* $x' \equiv x/\sqrt{\hat{z}(X)}$. Under reasonable regularity conditions, the distribution of x' tends to a Gaussian distribution as the estimate $\hat{z}(X)$ improves. As a demonstration of the ability of the model to account for image data, figure 2 shows the marginal histograms (in the log domain) and normal probability plots of x and x'. The histogram of the normalized coefficients is nearly Gaussian, and its corresponding normal probability plot lies nearly along a line. Additional examples may be found in previous publications.[17]

The power of the GSM model for the problem of denoising comes from the underlying Gaussian form. Consider the problem of estimating a single GSM coefficient $x = \sqrt{z}u$ that has been contaminated with additive Gaussian noise, w. The observation is written $y = \sqrt{z}u + w$. If the value of z were known, then the coefficient density would also be Gaussian distributed, and the optimal estimate would be the well-known linear (Wiener) solution:

$$\hat{x} = \frac{z\sigma_u^2}{z\sigma_u^2 + \sigma_w^2}\; y.$$

To this end, we focus on computing an estimate of the hidden multiplier z associated with each coefficient, from on observation of a surrounding cluster of coefficients contaminated by additive Gaussian noise.

3. ESTIMATION OF HIDDEN MULTIPLIERS

In order to use the GSM model for the application of image denoising, we need an estimator for the hidden multiplier z, and the covariance of the underlying Gaussian vector, C_u, in the presence of noise. Assume the wavelet coefficients are corrupted by additive noise:

$$Y = X + W,$$

where W is zero-mean Gaussian noise. Assuming that the noise is independent of the coefficients, X, the covariance of the noisy coefficient vector is:

$$\begin{aligned} C_y &= C_x + C_w \\ &= zC_u + C_w. \end{aligned}$$

Consider first the simple case where both U and W are decorrelated:

$$C_u = \sigma_u^2 I, \qquad C_w = \sigma_w^2 I,$$

with I the identity matrix. Without loss of generality, we can assume $\sigma_u = 1$, by absorbing the constant into the multiplier z. The maximum likelihood estimator is easily found by differentiating the log likelihood expression with respect to z and setting equal to zero:

$$\begin{aligned} \hat{z}(Y) &= \arg\max_z \left\{ -N \log(z + \sigma_w^2) - \frac{Y^T Y}{(z + \sigma_w^2)} \right\} \\ &= \frac{Y^T Y}{N} - \sigma_w^2. \end{aligned} \tag{1}$$

Thus, the variance of the coefficient at the center of the cluster is estimated from the average squared value of the neighbors. This concept has been used in a number of previous denoising approaches.[4,12,9,10]

Although wavelet coefficients of most photographic images are only weakly correlated, some specialized images do contain strong correlations[11]). In the case when the covariance matrix is not a multiple of the identity, the maximum likelihood estimate of z may be found by diagonalizing C_u.[11] Let Q be the orthogonal matrix containing the eigenvectors of C_u, and Λ be the diagonal matrix containing the associated eigenvalues λ_n, such that:

$$C_u = Q \Lambda Q^T.$$

The maximum likelihood estimate of z is the solution of:

$$\sum_{n=1}^{N} \frac{\lambda_n v_n^2}{(z\lambda_n + \sigma_w^2)^2} - \frac{\lambda_n}{z\lambda_n + \sigma_w^2} = 0, \tag{2}$$

where v_n are the components of vector $V = QY$, and the λ_n are the eigenvalues (diagonal elements of Λ).[11]

Finally, if the noise is correlated (for example, if the wavelet transform is not orthogonal), the maximum likelihood may be found by whitening W and then diagonalizing C_u. Let matrix S be the square root of C_w (i.e., the product of the eigenvector matrix and the square root of the diagonal eigenvalue matrix), such that $C_w = SS^T$. Now let Q and Λ contain the eigenvectors and eigenvalues of $S^{-1} C_u S^{-T}$, where S^{-T} indicates the transposed inverse of matrix S. Finally, the maximum likelihood estimate of z is the solution of:

$$\sum_{n=1}^{N} \frac{\lambda_n v_n^2}{(z\lambda_n + 1)^2} - \frac{\lambda_n}{z\lambda_n + 1} = 0, \tag{3}$$

where the v_n are the components of vector $V = QS^{-1}Y$.

In practice, we need also to estimate covariance C_u from a collection of noisy observations Y_k. If the number of observations K is large, the sample covariance may be approximated as

$$\frac{1}{K} \sum_{k=1}^{K} Y_k Y_k^T \approx C_y = C_w + \mu_z C_u,$$

where μ_z is the expected value of the multiplier variable z. Without loss of generality, this constant may be absorbed into the covariance matrix C_u. In our experiments, we assume C_w is known, and use the following estimate for C_u:

$$\hat{C}_u = \frac{1}{K} \sum_{k=1}^{K} Y_k Y_k^T - C_w. \tag{4}$$

4. DENOISING ALGORITHM

Given a method of estimating the hidden multiplier from the noisy data, we can now apply Gaussian scale mixture model to image denoising. For each coefficient in each subband of a multiresolution decomposition of a noisy image, we compute $\hat{z}(Y)$ from the vector of coefficients in the neighborhood, Y. Then the coefficient estimate is computed using the classical linear (Wiener) solution. In summary:

1. Perform multiscale decomposition of image corrupted by Gaussian noise.
2. For each subband (except the lowpass residual):
   a) Compute estimated covariance matrix $\hat{C}_u$ using equation (4).
   b) For each coefficient y (with surrounding neighborhood Y) in the subband:
      - Compute $\hat{z}(Y)$ numerically using equation (3).
      - Compute estimated variance of original coefficient x: $\hat{\sigma}_x^2 = \hat{z}(Y)\hat{\sigma}_u^2$, with $\hat{\sigma}_u^2$ the corresponding diagonal element of $\hat{C}_u$.
      - Replace noisy coefficient y by the linear estimate of the original coefficient $\hat{x} = \frac{\hat{\sigma_x}^2}{\hat{\sigma_x}^2 + \sigma_w^2} y$.
3. Invert the multiscale decomposition, reconstructing the denoised image from the estimated coefficients.

Note again that the algorithm estimates each multiplier z, and each coefficient x independently, even though the neighborhoods of adjacent coefficients are overlapping. The algorithm is currently somewhat computationally expensive, due to the numerical solution of equation (3) for each multiplier. The algorithm is closely related to that of Mihçak *et. al.*.[10] The main differences are the choice of basis (we use a steerable pyramid, as described in the next section), and the inclusion of signal and/or noise covariance matrices. The algorithm is also closely related to our previously developed algorithm.[11] In addition to the choice of basis, the main difference is that the previous algorithm was applied to non-overlapping blocks of coefficients.

5. NUMERICAL EXPERIMENTS

We performed a series of denoising experiments in order to test our local GSM method. In particular, we examined three different types of multiresolution decomposition: 1) orthogonal least-asymmetric wavelets,[23] 2) symmetric 9-tap QMF filters,[24] and 3) a four-orientation steerable pyramid.[25] The last of these is an overcomplete tight frame, in which the basis functions are polar separable in the Fourier domain. The primary advantage of this decomposition for denoising is that the subbands are sampled below the Nyquist rate, thus avoiding aliasing artifacts commonly seen with critically-sampled representations. We found that the symmetric QMF filters provided an improvement of 0.3-0.5 dB SNR over the asymmetric set, and the steerable pyramid provided an additional improvement of 0.4-0.6 dB. All results for GSM denoising reported below are computed using the steerable pyramid, with four orientations and four scales.

We also experimented with a number of different size and shape neighborhoods. For orthogonal wavelets and QMF filters the optimal choice was a rectangular neighborhood: 7 by 9 in horizontal detail subband, 9 by 7 in vertical detail subband, and 9 by 9 in diagonal detail subband. This indirectly confirms our intuition that the correlation structure is defined by local edges. For the steerable pyramid we found that a 7 by 7 neighborhood was sufficient for all subbands. Our previous work suggests that including coefficients of other subbands (i.e., corresponding to basis functions of other orientations and scales) should improve the estimates.[4,13]

image	thresholding	wiener2	simple GSM	covariance GSM
Lenna	27.96	28.13	30.57	30.60
Boats	27.06	27.26	29.19	29.32
Yogi	25.39	26.68	26.98	27.45
Barbara	24.06	24.28	25.49	25.21
Fingerprint	26.43	26.65	29.59	29.53
Einstein	28.61	28.98	30.90	31.37

Table 1. Comparison of denoising results. Values are peak-to-peak signal-to-noise ratio (PSNR) ($20\log_{10}(255/\sigma_{\text{error}})$), with added Gaussian noise of variance $\sigma_w = 25$ (PSNR = 20.17). All images are 512×512 pixels, except for "Einstein" which is 256×256.

noisy	thresholding	wiener2	simple GSM	covariance GSM
20.17 ($\sigma_w = 25$)	28.61	28.98	30.90	31.37
22.11 ($\sigma_w = 20$)	29.75	30.21	32.11	32.52
24.61 ($\sigma_w = 15$)	31.33	31.73	33.72	34.03
28.13 ($\sigma_w = 10$)	33.68	33.98	36.10	36.20

Table 2. Comparison of denoising results (PSNR, in dB) for the 256×256 "Einstein" image, with different amounts of additive Gaussian noise.

We compared our GSM method to two widely used denoising techniques. The first is the `wiener2` function implemented in MATLAB.[15] The algorithm computes a pixel-wise adaptive Wiener estimate based on the sample statistics of a local neighborhood of each pixel. In all experiments, we chose the neighborhood to maximize SNR. The second technique is *soft thresholding*, as developed by Donoho.[5] This method is based on an orthogonal wavelet decomposition using the Daubechies least-asymmetric 10-coefficient filters. In all experiments, the threshold is chosen for each subband to minimize mean squared error. We also used two variants of the GSM method. *Simple GSM* assumes uncorrelated coefficients ($C_u \propto I$), and uncorrelated noise of known variance ($C_w = \sigma_w^2 I$). *Covariance GSM* computes an estimate of C_u using equation (4), and assumes a known C_w (non-diagonal, due to the overcompleteness of the steerable pyramid).

Examples of denoising of the "Einstein" image are shown in Figure 3. The GSM approach appears both sharper and less noisy than the other two algorithms. In addition, both `wiener2` and `thresholding` approaches produce significant visual artifacts. A numerical comparison of the denoising results is given in Tables 5–5. One can see that the GSM approach significantly outperforms both `wiener2` and thresholding. In most examples, covariance GSM gives slightly better results than the simple GSM.

6. CONCLUSIONS

We have presented a denoising algorithm based on a Gaussian scale mixture model of local clusters of wavelet coefficients of photographic images. Our approach is extremely simple, due to the fact that individual wavelet coefficients, as well as their associated multipliers, are estimated independently. Despite this gross simplification, our denoising results are amongst the best reported in the literature.

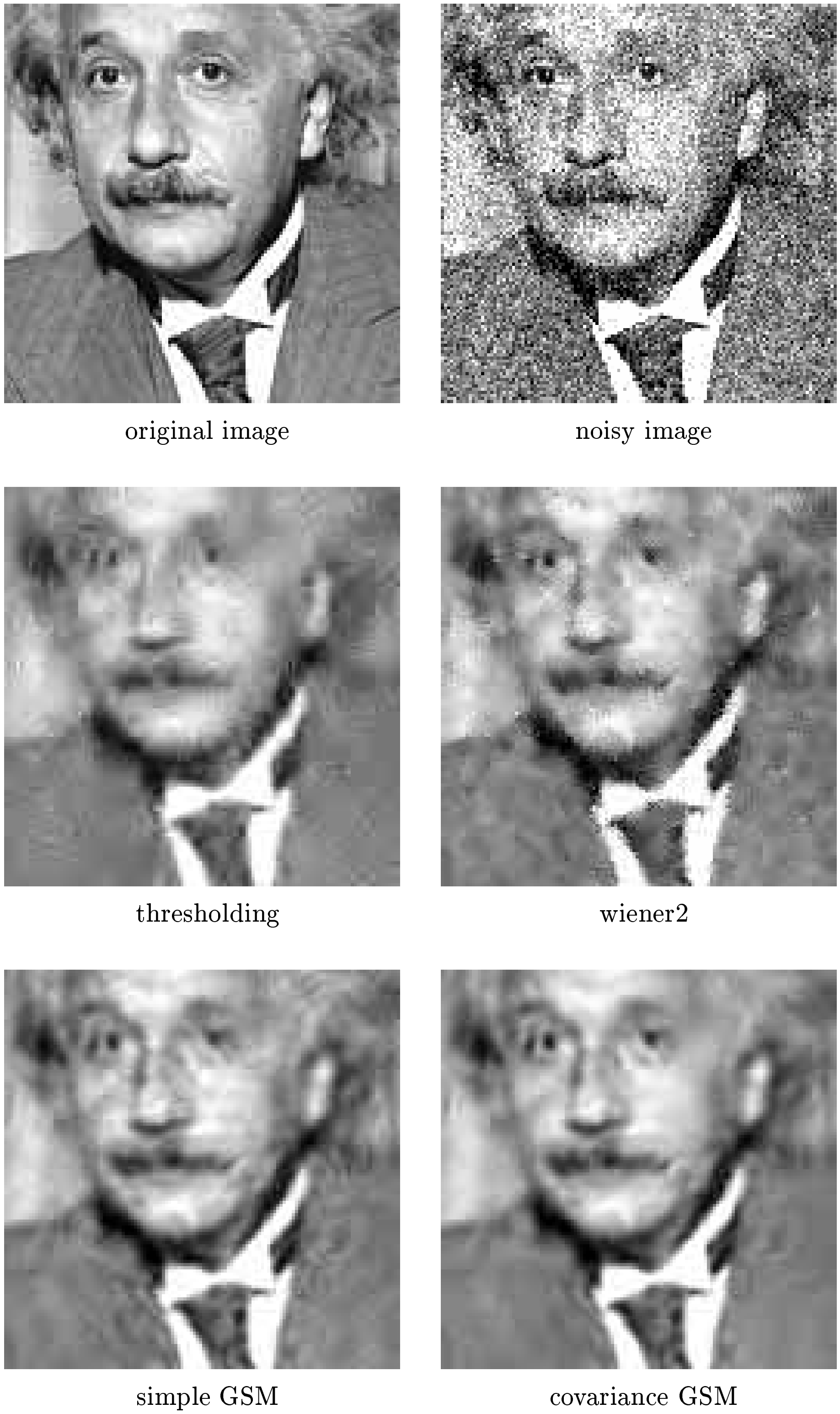


Figure 3. Results of denoising for the 256×256 "Einstein" image. Only a 128×128 cropped subregion of each image is shown. Original noise level was $\sigma_w = 25$ (see top row of Table 2).

noisy	thresholding	wiener2	simple GSM	covariance GSM
20.17 ($\sigma_w = 25$)	27.96	28.13	30.57	30.60
22.11 ($\sigma_w = 20$)	29.08	29.31	31.63	31.64
24.61 ($\sigma_w = 15$)	30.57	30.84	32.98	32.96
28.13 ($\sigma_w = 10$)	32.76	33.03	34.87	34.82

Table 3. Comparison of denoising results (PSNR, in dB) for the 523×523 "Lenna" image, with different amounts of additive Gaussian noise.

Our procedure is based on independent maximum likelihood estimates of the multiplier associated with each coefficient. We believe that performance can be somewhat improved by including a marginal prior on the multiplier variable z, as in the method of Mihçak *et. al.*.[10] Furthermore, our decision to estimate each multiplier independently greatly simplifies the algorithm, but clearly limits the performance. A more complete model should also consider the statistical relationship between adjacent multipliers, as in the model of Wainwright *et. al.*.[19,20] As such, our current algorithm provides a lower-bound for performance one can expect from a model with either a marginal or a global prior.

ACKNOWLEDGMENTS

Thanks to M. Wainwright for helpful comments on this manuscript. JP is partially supported by a fellowship from the Programa Nacional de Formacion de Personal Investigador (Spanish Government). EPS is partially supported by NSF CAREER grant MIP-9796040.

REFERENCES

1. D. J. Field, "Relations between the statistics of natural images and the response properties of cortical cells," *J. Opt. Soc. Am. A* **4**(12), pp. 2379–2394, 1987.
2. S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Pat. Anal. Mach. Intell.* **11**, pp. 674–693, July 1989.
3. C. Zetzsche, B. Wegmann, and E. Barth, "Nonlinear aspects of primary vision: Entropy reduction beyond decorrelation," in *Int'l Symposium, Society for Information Display*, vol. XXIV, pp. 933–936, 1993.
4. E. P. Simoncelli, "Statistical models for images: Compression, restoration and synthesis," in *31st Asilomar Conf on Signals, Systems and Computers*, pp. 673–678, IEEE Computer Society, (Pacific Grove, CA), November 1997. Available from http://www.cns.nyu.edu/~eero/publications.html.
5. D. Donoho, "Denoising by soft-thresholding," *IEEE Trans. Info. Theory* **43**, pp. 613–627, 1995.
6. E. P. Simoncelli and E. H. Adelson, "Noise removal via Bayesian wavelet coring," in *Third Int'l Conf on Image Proc*, vol. I, pp. 379–382, IEEE Sig Proc Society, (Lausanne), September 1996.
7. P. Moulin and J. Liu, "Analysis of multiresolution image denoising schemes using a generalized Gaussian and complexity priors," *IEEE Trans. Info. Theory* **45**, pp. 909–919, 1999.
8. 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), March 1997.
9. S. G. Chang, B. Yu, and M. Vetterli, "Spatially adaptive wavelet thresholding with context modeling for image denoising," in *Fifth IEEE Int'l Conf on Image Proc*, IEEE Computer Society, (Chicago), October 1998.
10. M. K. Mihçak, I. Kozintsev, K. Ramchandran, and P. Moulin, "Low–complexity image denoising based on statistical modeling of wavelet coefficients," *IEEE Trans. on Signal Processing* **6**, pp. 300–303, December 1999.
11. V. Strela, "Denoising via block Wiener filtering in wavelet domain," in *3rd European Congress of Mathematics*, Birkhäuser Verlag, (Barcelona), July 2000.

12. 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., ch. 18, pp. 291–308, Springer-Verlag, New York, Spring 1999. Lecture Notes in Statistics, vol. 141.
13. R. W. Buccigrossi and E. P. Simoncelli, "Image compression via joint statistical characterization in the wavelet domain," *IEEE Trans Image Proc* **8**, pp. 1688–1701, December 1999.
14. J. Portilla and E. Simoncelli, "Image denoising via adjustment of wavelet coefficient magnitude correlation," in *Seventh IEEE Int'l Conf on Image Proc*, IEEE Computer Society, (Vancouver), September 10-13 2000.
15. J. S. Lee, "Digital image enhancement and noise filtering by use of local statistics," *IEEE Pat. Anal. Mach. Intell.* **PAMI-2**, pp. 165–168, March 1980.
16. D. T. Kuan, A. A. Sawchuk, T. C. Strand, and P. Chavel, "Adaptive noise smoothing filter for images with signal-dependent noise," *IEEE Pat. Anal. Mach. Intell.* **PAMI-7**, pp. 165–177, March 1985.
17. 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., vol. 12, pp. 855–861, MIT Press, (Cambridge, MA), May 2000. Presented at Neural Information Processing Systems, Dec 1999.
18. M. S. Crouse, R. D. Nowak, and R. G. Baraniuk, "Wavelet-based statistical signal processing using hidden Markov models," *IEEE Trans. Signal Proc.* **46**, pp. 886–902, April 1998.
19. M. Wainwright, E. Simoncelli, and A. Willsky, "Random cascades on wavelet trees and their use in modeling and analyzing natural imagery," in *Proc SPIE, 45th Annual Meeting*, (San Diego), July 2000.
20. 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. Special issue on wavelet applications. To appear.
21. E. P. Simoncelli, "Modeling the joint statistics of images in the wavelet domain," in *Proc SPIE, 44th Annual Meeting*, vol. 3813, pp. 188–195, (Denver), July 1999.
22. D. Andrews and C. Mallows, "Scale mixtures of normal distributions," *J. Royal Stat. Soc.* **36**, pp. 99–, 1974.
23. I. Daubechies, "Ten lectures on wavelets," in *Regional Conference: Society for Industrial and Applied Mathematics*, (Philadelphia, PA), 1992.
24. E. P. Simoncelli and E. H. Adelson, "Subband transforms," in *Subband Image Coding*, J. W. Woods, ed., ch. 4, pp. 143–192, Kluwer Academic Publishers, Norwell, MA, 1990.
25. E. P. Simoncelli and W. T. Freeman, "The steerable pyramid: A flexible architecture for multi-scale derivative computation," in *Second Int'l Conf on Image Proc*, vol. III, pp. 444–447, IEEE Sig Proc Society, (Washington, DC), October 1995.