ADAPTIVE WIENER DENOISING USING A GAUSSIAN SCALE MIXTURE MODEL IN THE WAVELET DOMAIN

Javier Portilla[*], *Vasily Strela*[†], *Martin J. Wainwright*[‡], *Eero P. Simoncelli*[†]

[*]Dept. de Ciencia Computacional
Universidad de Granada
Spain
javier@decsai.ugr.es

[†]Ctr. for Neural Science, and
Courant Inst. Math. Sciences
New York University
{eero,vstrela}@cns.nyu.edu

[‡]Lab. Info. & Decision Systems
Dept. Elec. Eng. & Comp. Sci.
Mass. Inst. of Technology
mjwain@mit.edu

We describe a statistical model for images decomposed in an overcomplete wavelet pyramid. Each coefficient of the pyramid is modeled as the product of two independent random variables: an element of a Gaussian random field, and a hidden multiplier with a marginal log-normal prior. The latter modulates the local variance of the coefficients. We assume subband coefficients are contaminated with additive Gaussian noise of known covariance, and compute a MAP estimate of each multiplier variable based on observation of a local neighborhood of coefficients. Conditioned on this multiplier, we then estimate the subband coefficients with a local Wiener estimator. Unlike previous approaches, we (a) empirically motivate our choice for the prior on the multiplier; (b) use the full covariance of signal and noise in the estimation; (c) include adjacent scales in the conditioning neighborhood. To our knowledge, the results are the best in the literature, both visually and in terms of squared error.

1. INTRODUCTION

The images that we encounter in the real world have distinct features that make them very different from white noise. This enables humans to detect distortion in images, and to extract the remaining visually relevant information. The goal of image restoration is to release human observers from this task, by constructing a plausible original given the noisy observation. A prior probability model for uncorrupted images is of central importance for this application, as well as many others (e.g., image compression).

Modeling the statistics of natural images is a challenging task, because of the high dimensionality of the signal, and the complexity of statistical structures that are prevalent. Simplifying assumptions, such as homogeneity and locality are essential. In the past decade, it has become standard to decompose images with multi-scale band-pass oriented filters. These representations have been shown to decouple some high-order statistical features of natural images.

In this paper, we describe a stochastic model for local neighborhoods of coefficients of such a representation, in which the parameters are governed by a hidden random field. Specifically, local neighborhood of coefficients are modeled as the product of a Gaussian random vector and a hidden multiplier variable. We describe an efficient denoising method based on this model, and demonstrate the strength of the approach through numerical experiments.

2. IMAGE REPRESENTATION

Linear representations based on multi-scale band-pass oriented filters serve as a simple model for early processing in the human visual system. In addition, they are well-suited for representing basic properties of natural images, such as scale-invariance and the existence of locally oriented structures (i.e., edges). Due to a convergence of these qualitative properties, as well as an elegant mathematical framework, wavelets have emerged as the representation of choice for many image processing applications.

It has been observed, however, that overcomplete representations are superior to orthogonal wavelets for image denoising [e.g., 1, 2, 3, 4, 5]. The usual approach has been to use the same basis functions as in a critically-sampled orthogonal wavelet decomposition, but without decimating below the Nyquist rate [2]. However, once the critical sampling constraint has been eliminated, significant improvement comes from re-designing smoother filters that achieve rotation-invariance [4, 6]. For the current paper, we use a tight frame known as the "steerable pyramid" [1]. Like the undecimated wavelet transform, the subbands are translation-invariant. But the basis functions are symmetric and smooth, and (except for the high-pass and low-pass residuals) they are rotated and scaled versions of each other.

JP is supported in part by the Programa Nacional de FPI (Spanish Government). VS is on leave from Drexel University, and is supported by an AMS Centennial Fellowship. MW is supported by a NSERC-1967 Fellowship. EPS is supported by an NSF CAREER grant, an Alfred P. Sloan Fellowship, and the Howard Hughes Medical Institute.

3. STATISTICAL MODEL

3.1. Local GSMs in the wavelet domain

A random vector $\mathbf{x}$ is a Gaussian scale mixture (GSM) [7] if it can be expressed as the product of two independent random variables: $\mathbf{x} = \sqrt{z}\mathbf{u}$, where z is a positive scalar and $\mathbf{u}$ is a zero-mean Gaussian vector with covariance matrix C_u. The density of $\mathbf{x}$, denoted $p_x(\mathbf{x})$, is determined uniquely by $p_z(z)$ and C_u. In its one-dimensional version, the GSM class includes well-known symmetric heavy-tailed densities [8]. The key advantage of the GSM model is that the density of $\mathbf{x}$ is Gaussian when conditioned on z.

Marginal distributions of the wavelet coefficients of natural images are strongly non-Gaussian with heavy tails. In addition, the wavelet coefficients exhibit striking high-order joint statistical dependencies [9, 4]. Suppose $\mathbf{x}$ is a vector of wavelet coefficients at neighboring locations, orientations and scales (a generalized neighborhood). Empirically, it is found that the *amplitudes* of the components of $\mathbf{x}$ are correlated [9, 4]. This is partially a consequence of the fluctuation in local contrast at different locations: large-amplitude coefficients tend to be near each other. But even after normalization of the local contrast in a natural image, this effect persists. The reason is the existence of oriented local structures (edges, lines, etc.), that couple the variance of the wavelet coefficients at adjacent scales, orientations and locations. A multi-scale and multi-orientation representation is thus essential for capturing the statistical dependencies created by these structures.

We have previously shown that a GSM model can account for both the marginal and pairwise joint empirical distributions of wavelet coefficients [8]. Local GSM models, which characterize a neighborhood of coefficients with a single multiplier, have been used successfully for denoising [11, 12, 6] (see [10, 13] for global GSM models that include dependencies amongst multipliers). In this paper, we extend these results by developing a more realistic prior model for the hidden multiplier, by including a coefficient from a coarser scale in the GSM vector $\mathbf{x}$, and by using a full covariance description of both the Gaussian vector $\mathbf{u}$ and the noise.

3.2. Prior distribution of the multiplier

We use a non-parametric technique to determine the shape of the marginal prior density of the multiplier $\sqrt{z}$ from coefficient histograms of a set of uncorrupted images. Consider the marginal version of the GSM model. Taking absolute values and logarithms gives: $\log|x| = \frac{1}{2}\log z + \log|u|$. Because z and u are independent the density of $\log|x|$ may be written as a convolution:

$$p_{\log|x|}(\log|x|) = p_{\log|u|}(\log|u|) * 2p_{\log z}(2\log z).$$

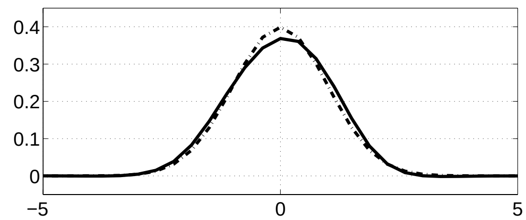

Fig. 1. Empirically measured distribution $p_{\log z}(\log z)$, compared to a Gaussian (dashed line).

Without loss of generality, we assume u has same variance as x. Thus, $p_{\log z}(\log z)$ can be estimated by deconvolving the empirically estimated density $p_{\log|x|}(\log|x|)$ (note: due to the estimation noise in the histogram of $\log|x|$, a regularization technique is required).

In Fig. 1 we show the resulting prototype for $p_{\log z}(\log z)$, obtained from an average over the histograms of 306 subbands of 18 test images (histograms are normalized for mean and variance before averaging). The result is remarkably symmetric, and close to Gaussian (dashed line). Thus, we choose a log-normal model for $p_z(z)$:

$$p_z(z) = \frac{\exp\left(-(\log z - \mu_l)^2/(2\sigma_l^2)\right)}{z(2\pi\sigma_l^2)^{1/2}}. \qquad (1)$$

Note that the density of $\sqrt{z}$ will also be log-normal.

4. DENOISING

We follow a now standard approach for denoising in the wavelet domain. We begin by decomposing the image into 18 pyramid subbands (4 orientations at each of 4 scales, plus high-pass and low-pass residuals). For each band (except the low-pass) we apply the denoising method explained below. Although the subbands are processed sequentially, they are *not* processed independently, since the conditioning neighborhoods include coefficients from coarser scales. The denoised image is computed by inverting the pyramid transform.

Suppose an image is corrupted by independent additive Gaussian noise. A vector $\mathbf{y}$ corresponding to a generalized neighborhood of observed noisy coefficients can be expressed as:

$$\mathbf{y} = \sqrt{z}\mathbf{u} + \mathbf{w},$$

where $\mathbf{u}$ and $\mathbf{w}$ are zero-mean Gaussian vectors with covariance matrices C_u and C_w. We wish to estimate $\mathbf{x} = \sqrt{z}\mathbf{u}$, assuming a log-normal prior for z. Although optimal estimation is difficult, a good solution arises naturally from the structure of the GSM model: (1) estimate z, (2) conditioned on z, estimate $\mathbf{x}$. Because $\mathbf{y}$ is Gaussian conditioned on z, the second step may be computed optimally as a linear (Wiener) estimate [e.g., 14, 4, 11, 12, 6, 5].

4.1. Parameter estimation

The parameters of our GSM model consist of the covariances $\{C_u, C_w\}$, and the mean and variance of the (log) multiplier distribution, $\{\mu_l, \sigma_l^2\}$. Since $\mathbf{u}$ and $\mathbf{w}$ are independent, the covariance matrix of $\mathbf{y}$ conditioned on z is $C_{y|z} = zC_u + C_w$. Taking the expectation over z yields $C_y = \mathcal{E}_z\{z\}C_u + C_w$. Without loss of generality we set $\mathcal{E}_z\{z\} = 1$, obtaining $C_u = C_y - C_w$. C_y is estimated from each noisy subband and C_w is computed from the power spectral density of the noise (flat, if white) and the pyramid basis functions.

For the estimate of μ_l and σ_l^2, we use the method of moments. First we estimate σ_z^2 from the expression $\mathcal{E}\{y^4\} = \mathcal{E}\{(\sqrt{z}u + w)^4\}$, where $\mathcal{E}\{y^4\}$ is estimated from the data: $\sigma_z^2 = (\mathcal{E}\{y^4\}/3 - \sigma_y^4)/(\sigma_y^2 - \sigma_w^2)^2$. Next we use equation (1) to express the mean and variance of z as functions of μ_l and σ_l^2, and equate them to 1 and σ_z^2, respectively. Solving these equations yields: $\mu_l = -\log(\sigma_z^2 + 1)/2$ and $\sigma_l^2 = \log(\sigma_z^2 + 1)$

4.2. Multiplier estimation

For simplicity, each multiplier is estimated independently, ignoring the overlap of the neighborhoods. As in [11, 6], we compute a maximum a posteriori (MAP) estimate of z:

$$\hat{z} = \arg\max_z \{p_{z|\mathbf{y}}(z|\mathbf{y})\} = \arg\max_z \{p_{\mathbf{y}|z}(\mathbf{y}|z)p_z(z)\},$$

where

$$p_{\mathbf{y}|z}(\mathbf{y}|z) = \frac{\exp\left(-\mathbf{x}^T(zC_u + C_w)^{-1}\mathbf{x}/2\right)}{((2\pi)^N|zC_u + C_w|)^{1/2}}, \qquad (2)$$

and N is the number of neighbors. Multiplying the right hand sides of equations (1) and (2), taking logarithms, differentiating the result with respect to z and equating to 0, yields:

$$\frac{\log z - \mu_l + \sigma_l^2}{z\sigma_l^2} + \sum_{n=1}^{N} \frac{z - z_n}{(z + \lambda_n^{-1})^2} = 0, \qquad (3)$$

where $z_n = (v_n^2 - 1)/\lambda_n$, v_n are the components of the vector $V = Q^T S^{-1}\mathbf{y}$, $S = C_w^{1/2}$ (i.e., $C_w = SS^T$), λ_n are the diagonal components of Λ, and (Q, Λ) are the eigenvector and eigenvalue matrices of $S^{-1}C_u S^{-T}$, respectively. Note that S, Q, and Λ are computed once for each subband. We apply the bisection method to solve this equation numerically, for each coefficient in the pyramid.

4.3. Coefficient estimation

Conditioned on z, $\mathbf{y}$ is Gaussian and may be estimated using the Wiener filter (optimal both in maximum likelihood and least square senses):

$$\hat{\mathbf{x}} = \hat{z}C_u(\hat{z}C_u + C_w)^{-1}\mathbf{y},$$

which in terms of S, Λ, Q, V becomes

$$\hat{\mathbf{x}} = SQ(I + \hat{z}^{-1}\Lambda^{-1})^{-1}V.$$

This expression gives an estimate for the whole neighborhood, but we use it only for the central coefficient. Thus, our final estimate is $\hat{x} = \sum_{n=1}^{N} m_{c,n}v_n/(1 + \hat{z}^{-1}\lambda_n^{-1})$, where $m_{i,j}$ represents an element of $M = SQ$, and c is the index of the central coefficient.

4.4. Implementation

We have explored many combinations of neighbor positions, scales and orientations for inclusion in the generalized neighborhood. Results depend on the noise level, but for moderate to high levels of noise, a 5×5 square region around each coefficient, together with the coefficient at the same location and orientation at the next coarser scale (the *parent*), is roughly optimal. We denote this generalized neighborhood $5 \times 5 + p$. Another practical issue is convolution boundary handling: we have used mirror-symmetric extension.

Our method is moderately demanding in terms of computation. Our Matlab implementation takes 3.7 minutes to process a 256×256 image on a 900 Mhz Pentium III, and 12.8 minutes for a 512×512 image. These times are reduced by roughly a factor of two with a $3 \times 3 + p$ neighborhood, at the expense of a small decrease in performance (see Table 1).

5. RESULTS

Figure 2 compares the results of our numerical experiments with the best available published results [3, 11, 5]. Two common test images and five noise levels are used. We find the improvement impressive. Figure 3 shows visual result of denoising of the *Barbara* image, compared with the results reported in [5]. Our method is seen to provide better preservation of edges and other details. Also, the separation of diagonal orientations in the steerable pyramid allows more selective removal of the noise in diagonally oriented image regions (see lines on the left side of the image).

Table 1 shows the decrease in performance that results when various features of our model are removed. The first two columns examine reduced neighborhoods. Removal of the parent (see column *Prnt*) has roughly the same impact as removing the outer ring of 16 spatial neighbors (3×3). This is most significant for higher noise levels. The next two columns examine simplifications of the probability model. Eliminating the multiplier prior model (i.e., using a Maximum Likelihood estimate) leads to noticeable degradation for higher noise levels (*Prior*). The importance of using the full covariance of both noise and signal (as opposed to just the variance) is largely due to the strong correlation of coefficients in the steerable pyramid (*Cov*). The use of periodic

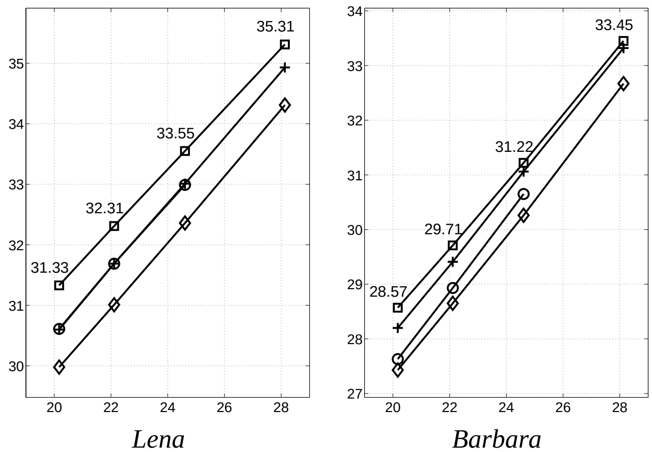

Fig. 2. Denoising performance. Output PSNR vs. input PSNR, in dB, for two images at four noise levels ($\sigma_w = \{25, 20, 15, 10\}$). Comparison with three state-of-the-art methods: diamonds [11]; circles [3]; crosses [5]; squares indicate our results, with associated PSNR values.

σ_w / PSNR	*Prnt*	3×3	*Prior*	*Cov*	*Bdry*
10 / 28.13	0.03	0.02	0.01	0.28	0.17
25 / 20.17	0.09	0.09	0.15	0.40	0.22
40 / 16.09	0.11	0.16	0.11	0.47	0.25

Table 1. Reduction in denoising performance (dB) due to removal of model components (see text). Results averaged over 5 images: *Lena*, *Barbara*, *Boats*, *Yogi*, *Einstein*.

convolution (as opposed to symmetric extension) leads to a surprisingly large drop in performance (*Bdry*). Finally, although not shown here, we also find that the impact of both the prior and the parent is dramatically larger when used with a smaller (3×3+p) neighborhood.

Fig. 3. Left to right and top to bottom: Cropped original *Barbara* image; noisy image ($\sigma_w = 25$, 20.17dB PSNR); result from [5] (28.20dB); our result (28.57dB).

6. REFERENCES

[1] 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,

[2] R R Coifman and D L Donoho, “Translation–invariant de–noising,” in *Wavelets and statistics*, A Antoniadis and G Oppenheim, Eds. Springer-Verlag, San Diego, 1995.

[3] S G Chang, Bin Yu, and Martin Vetterli, “Spatially adaptive wavelet thresholding with context modeling for image denoising,” in *IEEE Int’l Conf on Image Proc*, October 1998.

[4] 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., v. 141, ch. 18, pp. 291–308. Springer-Verlag, New York, 1999.

[5] Xin Li and Michael T. Orchard, “Spatially adaptive image denoising under overcomplete expansion,” in *IEEE Int’l Conf on Image Proc*, September 2000.

[6] V Strela, J Portilla, and E Simoncelli, “Image denoising using a local Gaussian scale mixture model in the wavelet domain,” in *Proc SPIE, 45th Annual Meeting*, San Diego, July 2000.

[7] D Andrews and C Mallows, “Scale mixtures of normal distributions,” *J. Royal Stat. Soc.*, vol. 36, pp. 99–, 1974.

[8] 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., May 2000, vol. 12, pp. 855–861.

[9] E P Simoncelli, “Statistical models for images: Compression, restoration and synthesis,” in *Proc 31st Asilomar Conf on Signals, Systems and Computers*, Pacific Grove, CA, November 1997, pp. 673–678, Available at http://www.cns.nyu.edu/∼eero/publications.html.

[10] 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.

[11] 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*, vol. 6, no. 12, pp. 300–303, December 1999.

[12] V Strela, “Denoising via block Wiener filtering in wavelet domain,” in *3rd European Congress of Mathematics*, Barcelona, July 2000, Birkhäuser Verlag.

[13] M J Wainwright, E P Simoncelli, and A S Willsky, “Random cascades on wavelet trees and their use in modeling and analyzing natural imagery,” *Applied and Computational Harmonic Analysis*, June 2001. In press.

[14] J S Lee, “Digital image enhancement and noise filtering by use of local statistics,” *IEEE Pat. Anal. Mach. Intell.*, vol. PAMI-2, pp. 165–168, March 1980.