Low-complexity image denoising based on statistical modeling of wavelet coefficients

M. Kıvanç Mıhçak, Igor Kozintsev, Kannan Ramchandran and Pierre Moulin

University of Illinois at Urbana-Champaign
Beckman Institute and ECE Department
405 N. Mathews Ave., Urbana, IL 61801
Email: {*mihcak,igor,kannan,moulin*}*@ifp.uiuc.edu*, phone: (217) 244-1089, fax:(217) 244 8371

Contact Author: M. K. Mıhçak, tel: (217) 244-1089, fax: (217) 244-8371,
Email: *mihcak@ifp.uiuc.edu*

March 22, 2004

EDICS Numbers: SPL.IP.1.4, SPL.SP.3.5, SPL.SP.3.6

Abstract

We introduce a simple spatially adaptive statistical model for wavelet image coefficients and apply it to image denoising. Our model is inspired by a recent wavelet image compression algorithm, the Estimation Quantization coder. We model wavelet image coefficients as zero-mean Gaussian random variables with high local correlation. We assume a marginal prior distribution on wavelet coefficients variances and estimate them using an approximate Maximum A Posteriori Probability rule. Then we apply an approximate Minimum Mean Squared Error estimation procedure to restore the noisy wavelet image coefficients. Despite the simplicity of our method, both in its concept and *implementation*, our denoising results are among the best reported in the literature.

Keywords

Image denoising, wavelets, statistical modeling, parameter estimation.

1 Introduction and motivation

Accurate image modeling, whether done explicitly or implicitly, is a critical component of many image processing tasks. In [?], a simple yet effective statistical *spatially-adaptive* wavelet image model was developed and formed the basis of the state-of-the-art Estimation-Quantization (EQ) compression algorithm. In this work, we develop a closely related model for image wavelet coefficients and apply it to denoising of images corrupted by additive white Gaussian noise (AWGN). Our new model significantly reduces the computational burden of an earlier version of our scheme in [?], yet produces comparable results in terms of mean-squared error (MSE) and perceptual image quality. The key ingredient of our new algorithm is the use of simple but efficient *spatial adaptation techniques*. This paper does not attempt to investigate the theoretical properties of the proposed models and algorithms in general settings. Our primary goal is to demonstrate the importance of accurate modeling for image denoising problems.

There is a close relationship between image compression and image denoising. In fact, the use of lossy data compression itself was proposed for denoising with the intuition that a "typical correlated signal is compressible but noise is not " [?] . This principle is also apparent in more theoretically motivated methods such as Minimum Description Length (MDL) and complexity regularized denoising [?, ?, ?]. In this paper, we further exploit the relationship between compression and denoising by using state-of-the-art image models developed for compression.

A wide class of image processing algorithms is based on the Discrete Wavelet transform. The transform coefficients within subbands can be modeled as independent identically distributed (i.i.d.) random variables with Generalized Gaussian (GG) distribution [?]. More sophisticated models for image compression recognize the existence of significant spatial dependencies in the transform coefficients, and try to describe these dependencies using data structures such as zero-trees [?]. The high performance of the zero-tree based image coders has led several researchers to develop similar methods for image denoising. A Hidden Markov model based on wavelet trees was proposed for denoising of 1-D signals in [?] and extended to image denoising in [?].

Zero-trees are not the only way to account for dependencies between wavelet coefficients. The EQ algorithm, which offers excellent compression performance, uses a very simple and efficient local model [?]. In this work, we modify this model for the purpose of image denoising, and demonstrate the benefits of this approach. A related model, that accounts for local dependencies, was independently proposed in [?] and its effectiveness was verified by various experimental results. In [?], another related adaptive model was used to perform image denoising via wavelet thresholding using context modeling of the *global* coefficients histogram. In our work, we take an opposite approach which exploits the *local* structure of wavelet image coefficients. Also, we use linear Minimum Mean Squared Error(MMSE)-like estimation instead of coefficient thresholding.

2 Proposed Approach

2.1 Stochastic model for wavelet coefficients

We model image wavelet coefficients as a realization of a doubly stochastic process. Specifically, the wavelet coefficients are assumed to be conditionally independent zero-mean Gaussian random variables, given their variances. These variances are modeled as identically distributed, highly correlated random variables. For estimation purposes, we approximate wavelet coefficients as locally i.i.d. Our proposed model differs from the one in [?] in the following two respects. First, we put a stochastic prior on the local variances rather than considering them as unknown deterministic parameters. Second, we model the wavelet coefficients as *conditionally independent Gaussian* random variables rather than GG variables as in [?]. To motivate this model, see Fig. **??**. The left panel shows the histogram of the original high-band coefficients of Lena image from the first scale, obtained by employing Daubechies-8 wavelets [?]. The right panel shows those coefficients normalized by their estimated standard deviations. Observe that the normalized histogram is well approximated by a zero-mean, unit-variance, Gaussian probability density function (p.d.f.).

In this paper, we assume that image pixels are corrupted by AWGN with known variance σ_n^2. Let $X(k)$ represent the orthonormal wavelet coefficients of the "clean" image. The wavelet coefficients of the noisy image are given by $Y(k) = X(k) + n(k)$, where $n(k)$ is AWGN due to orthonormality of the chosen wavelet transform.

Our proposed denoising algorithm operates in two steps. Initially, we perform an approximate MAP estimation of the variance $\sigma^2(k)$ for each coefficient, using the observed noisy data in a local neighborhood and a prior model for $\sigma^2(k)$. The estimate $\hat{\sigma}^2(k)$ is then substituted for $\sigma^2(k)$ in the expression for the MMSE estimator of $X(k)$. Both steps are summarized in Fig. **??**(a) and are described in more detail below.

2.2 Denoising algorithm

Given $\sigma^2(k)$, the wavelet coefficients $X(k)$ are independent Gaussian variables, so the MMSE estimator for $X(k)$ is given by $\hat{X}(k) = \frac{\sigma^2(k)}{\sigma^2(k)+\sigma_n{}^2} Y(k)$. We emphasize that this assumes $\sigma^2(k)$ is deterministic and known. But in fact $\sigma^2(k)$ is not known, so we construct a linear MMSE-like estimator $\hat{X}(k) = \frac{\hat{\sigma}^2(k)}{\hat{\sigma}^2(k)+\sigma_n{}^2} Y(k)$, where $\hat{\sigma}^2(k)$ is an estimate for $\sigma^2(k)$.

Our results indicate that the performance of the proposed approximate MMSE predictor is dependent, to some high extent, on the quality of the estimator $\hat{\sigma}^2(k)$. Generally, this relation is unknown and complicated; but it is likely that a better estimator for the data variance yields a better estimate for the data as well.

2.3 Estimation of the underlying variance field

The estimation of the variance field $\sigma^2(k)$ is the crux of the proposed denoising algorithm. For each data point $Y(k)$, an estimate of $\sigma^2(k)$ is formed based on a local neighborhood $\mathcal{N}(k)$. We use a square window $\mathcal{N}(k)$ centered at $Y(k)$. Assuming the correlation between variances of neighboring coefficients is high, we have $\sigma^2(j) \approx \sigma^2(k) \quad$ for all $j \in \mathcal{N}(k)$. Then we compute an approximate Maximum Likelihood (ML) estimator:

$$\hat{\sigma}^2(k) = arg \max_{\sigma^2 \geq 0} \prod_{j \in \mathcal{N}(k)} P(Y(j)|\sigma^2) = \max\left(0, \frac{1}{M} \sum_{j \in \mathcal{N}(k)} Y^2(j) \ - \sigma_n^2\right), \tag{1}$$

where $P(\cdot|\sigma^2)$ is the Gaussian distribution with zero mean and variance $\sigma^2 + \sigma_n^2$, and M is the number of coefficients in $\mathcal{N}(k)$. Now, assume a prior marginal distribution $f_\sigma\left(\sigma^2\right)$ for each $\sigma^2(k)$ is available. Then we obtain an approximate MAP estimator for $\sigma^2(k)$ as

$$\hat{\sigma}^2(k) = arg \max_{\sigma^2 \geq 0} \left[\prod_{j \in \mathcal{N}(k)} P(Y(j)|\sigma^2)\right] f_\sigma(\sigma^2).$$

In Fig. **??**(b), we plot a histogram of the estimated local variances using (**??**) with a 7×7 window for a typical high-pass subband of an image (solid line). The exponential prior $f_\sigma(\sigma^2) = \lambda e^{-\lambda\sigma^2}$ shown by the dash-dotted line in Fig. **??**(b), is a reasonable candidate to fit the original histogram. The approximate MAP estimate for $\sigma^2(k)$ using an exponential prior is given by:

$$\hat{\sigma}^2(k) = \max\left(0, \frac{M}{4\lambda}\left[-1 + \sqrt{1 + \frac{8\lambda}{M^2} \sum_{j \in \mathcal{N}(k)} Y^2(j)}\right] - \sigma_n^2\right). \tag{2}$$

In our experiments we indeed found out that (**??**) improves the denoising performance over (**??**). Using the exponential prior requires the estimation of only one additional parameter λ per image wavelet subband. One might expect that more accurate modeling of the histogram in Fig. **??**(b) would yield further improvements in denoising performance. For instance, we fitted the histogram with different exponentials in three non-overlapping regions (dashed line). In this case, the approximate MAP estimator still admits a simple closed-form expression. However, we experimentally verified that it does not produce better image denoising results.

3 Results and discussion

We tested our algorithm on a number of images, but only report results for *Lena* and *Barbara*. We used an orthogonal wavelet transform with five levels of decomposition and Daubechies' length-8 wavelet. Centered square-shaped windows of sizes 3×3, 5×5 and 7×7 were employed to find different estimates for $\sigma^2(k)$. The parameter λ of the prior f_σ was set equal to the inverse of the standard deviation of wavelet coefficients that were initially denoised by using (**??**) and linear MMSE-like estimation.

We compared five different denoising methods. The PSNR results are shown in Table **??**. The first method is the hard–thresholding of wavelet coefficients using a constant threshold for all subbands, calculated according to [**?**]. The second method is MATLAB's image denoising algorithm *wiener2*. The third method uses spatially adaptive wavelet thresholding [**?**]. We included only the results from [**?**] which were obtained by using an orthogonal wavelet transform since this is equivalent to our setup.

Our results are presented for two different methods. First, we treated the variances as deterministic quantities and computed approximate ML estimates. We call the resulting method **LAWML** (Locally Adaptive Window-based denoising using ML). The second method uses an exponential distribution as a prior for the underlying variance field. Based on this model, we compute approximate MAP estimates of the variances. We call this method **LAWMAP** (Locally Adaptive Window-based denoising using MAP).

In this work, we confined ourselves to square–shaped neighborhoods with fixed size, for simplicity. In general, it would be desirable to automatically select both the size and the shape of the neighborhood region. But clearly, this would introduce additional difficulties. The selection of the window size suggests a trade-off which has been discussed in detail in [**?**]. The flexibility of our proposed method lends itself to the usage of different shaped neighborhoods for each coefficient. This could be implemented by using edge- and shape-adapted windows. Such an adaptation is likely to further improve denoising performance, see [**?**] for an example.

References

[1] S. LoPresto, K. Ramchandran, and M. T. Orchard, "Image coding based on mixture modeling of wavelet coefficients and a fast estimation–quantization framework," in *Proc. IEEE Data Compression Conf.*, Snowbird, Utah, pp. 221–230, 1997.

[2] M. K. Mıhçak, I. Kozintsev, and K. Ramchandran, “Spatially adaptive statistical modeling of wavelet image coefficients and its application to denoising,” in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Proc.*, vol. 6, Phoenix, AZ, pp. 3253–3256, March 1999.

[3] B. Natarajan, “Filtering random noise from deterministic signals via data compression,” *IEEE Trans. on Signal Processing*, vol. 43, no. 10, pp. 2595–2605, Nov 1995.

[4] N. Saito, “Simultaneous noise suppression and signal compression using a library of orthonormal bases and the MDL criterion,” in *Wavelets in Geophysics* (E. Foufoula-Georgiou and P. Kumar, eds.), pp. 299–324, Academic Press, New York, 1994.

[5] J. Liu and P. Moulin, “Complexity-regularized image denoising,” in *Proc. IEEE Int. Conf. Image Proc.*, vol. 2, Santa Barbara, CA, pp. 370–373, 1997.

[6] P. Moulin and J. Liu, “Analysis of multiresolution image denoising schemes using generalized - gaussian and complexity priors,” *IEEE Trans. Info. Theory, Special Issue on Multiscale Analysis*, vol. 45, no. 3, pp. 909–919, April 1999.

[7] S. Mallat, “A theory for multiresolution signal decomposition: the wavelet representation,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 7, pp. 674–693, July 1989.

[8] J. M. Shapiro, “Embedded image coding using zerotrees of wavelet coefficients,” *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3445–3462, December 1993.

[9] M. Crouse, R. Nowak, and R. Baraniuk, “Wavelet-based statistical signal processing using hidden Markov models,” *IEEE Trans. on Signal Processing*, vol. 42, no. 4, pp. 886–902, April 1998.

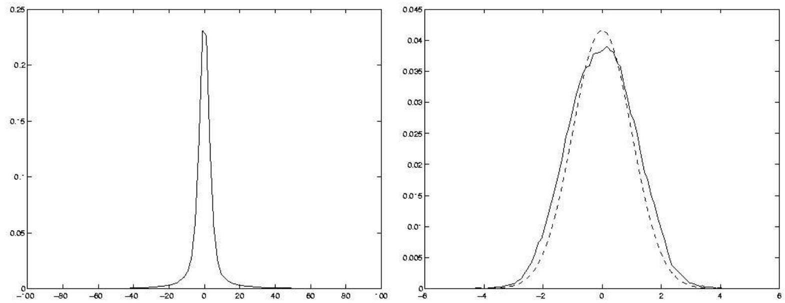
[10] J. Romberg, H. Choi, and R. Baraniuk, “Bayesian tree-structured image modeling using wavelet-domain hidden Markov models,” in *SPIE Technical Conference on Mathematical Modeling, Bayesian Estimation, and Inverse Problem*, Denver, Colorado, July 1999.

[11] E. P. Simoncelli, “Modeling the Joint Statistics of Images in the Wavelet Domain,” in *SPIE Conf. 3813 on Wavelet Applications in Signal and Image Processing VII*, Denver, Colorado, July 1999.

[12] S. G. Chang, B. Yu, and M. Vetterli, “Spatially adaptive wavelet thresholding with context modeling for image denoising,” *Submitted to IEEE Trans. on Image Processing*, 1998. Available from http://gabor.eecs.berkeley.edu/~grchang/publications.html.

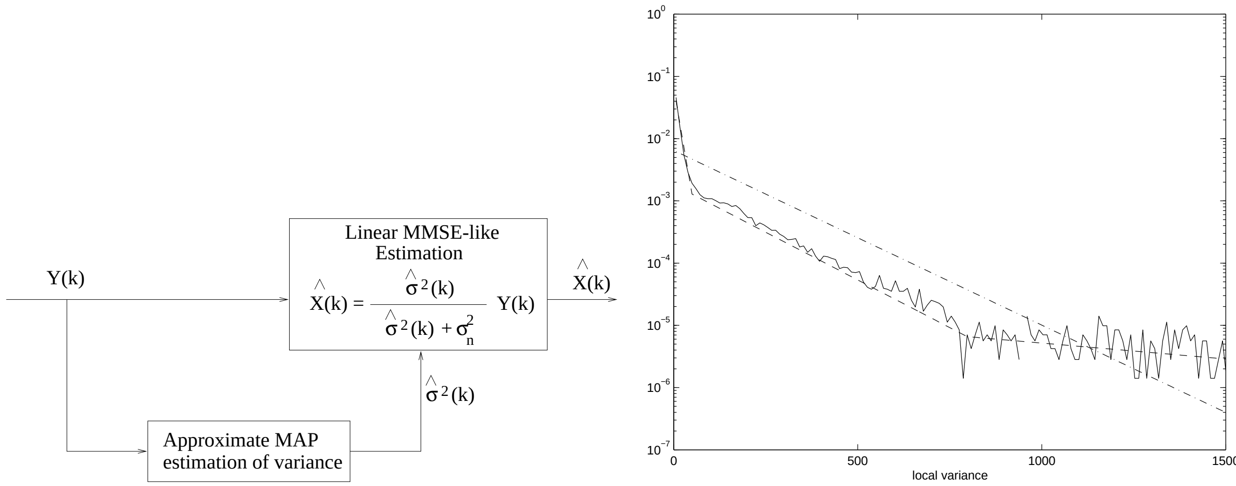
[13] M. Vetterli and J. Kovacevic, *Wavelets and Subband Coding*. Englewood Cliffs: Prentice-Hall, 1995.

[14] D. L. Donoho and I. M. Johnstone, “Ideal spatial adaptation by wavelet shrinkage,” *Biometrika*, vol. 81, pp. 425–455, 1994.



(a) (b)

Figure 1: (a) - Histogram of the high-band wavelet coefficients of *Lena*. (b) Solid line: histogram of the same coefficients scaled by estimated local standard deviations. Dashed line: unit-variance, zero-mean Gaussian p.d.f.



(a) (b)

Figure 2: (a) Block-diagram of the denoising algorithm. For each observed noisy coefficient $Y(k)$ we form an approximate MAP estimate $\hat{\sigma}^2(k)$ of the variance of $X(k)$ based on a local neighborhood and on the prior f_σ. The estimate $\hat{\sigma}^2(k)$ is then used for linear MMSE-like estimation of $X(k)$. (b) Histogram of the estimated local variance of the coefficients (solid line) in wavelet image subband approximated using a single exponential prior (dash-dotted line) and a mixture of exponentials that consists of three single exponentials in three non-overlapping regions (dashed line).

Table 1: PSNR (dB) results for several denoising methods

	Noise Standard Deviation σ_n			
	10	15	20	25
	LENA			
Donoho's hard thresholding [?]	30.34	28.52	27.24	26.34
Spatial local Wiener (MATLAB)	32.98	30.44	28.52	26.95
SAWT [?]	—	31.83	30.49	29.50
3x3 LAWML	33.72	31.37	29.63	28.22
3x3 LAWMAP	34.25	32.33	31.00	29.96
5x5 LAWML	34.13	31.99	30.46	29.24
5x5 LAWMAP	34.31	32.36	31.01	29.98
7x7 LAWML	34.17	32.10	30.65	29.52
7x7 LAWMAP	34.24	32.27	30.92	29.90
	BARBARA			
Donoho's hard thresholding [?]	27.29	25.01	23.65	22.83
Spatial local Wiener (MATLAB)	31.35	28.58	26.67	25.19
SAWT [?]	—	29.19	27.65	26.52
3x3 LAWML	32.32	29.72	27.93	26.53
3x3 LAWMAP	32.46	30.03	28.39	27.21
5x5 LAWML	32.54	30.09	28.43	27.15
5x5 LAWMAP	32.57	30.19	28.59	27.42
7x7 LAWML	32.49	30.09	28.49	27.28
7x7 LAWMAP	32.51	30.13	28.57	27.40