Two-Step Variance-Adaptive Image Denoising

Lahouari Ghouti and Ahmed Bouridane

Abstract— **In this paper, we describe a two-step variance-adaptive method for image denoising based on a statistical model of the coefficients of balanced multiwavelet transform. The model is derived in a statistical framework from a recent successful scheme developed in the seemingly unrelated front of lossy image compression. Clusters of multiwavelet coefficients are modeled as zero-mean Gaussian random variables with high local correlation. In the adopted framework, we use marginal prior distribution on the variances of the multiwavelet coefficients. Then, estimates of the local variances are used to restore the noisy multiwavelet coefficients based on a minimum mean square error estimation (MMSE) procedure. Experimental results, using images contaminated with additive white Gaussian noise, show that the proposed method outperforms most of the denoising schemes reported in the literature. In this paper, the performance comparison is restricted to non-redundant multiresolution representations.**

I. Introduction

During acquisition or transmission, an image gets often contaminated by noise. Therefore, it is desirable to derive an estimate of the original image through denoising using some estimation procedure. Recently, there has been an emergence of denoising schemes based on statistical modeling of image statistics. However, an accurate implicit or explicit modeling of the statistics of natural images is a critical component of many image processing tasks and in the same time a challenging task, partly because of the high dimensionality of the signal [1]. To circumvent the dimensionality problem, most of the available statistical models rely on two basic assumptions: 1) spatial *homogeneity* where it is assumed that the distribution of values in a neighborhood is the same for all such neighborhoods, regardless of absolute spatial position [1]; 2) *local adaptivity* where the probability structure can be defined *locally*. These two assumptions lead to a Markov random field model that can be further simplified by assuming Gaussian distributions [1]. Over the past decade, it has become a standard to boost the power of statistical image models by transforming the signal from the pixel domain to a multiresolution representation. This class of image processing algorithms, loosely referred to as wavelet-based algorithms, are characterized by a *decoupling* property where the high-order statistical features of natural images are efficiently decoupled [1], [2]. In recent years, more sophisticated models for homogeneous local probability have been developed for images in multiscale representations. These models investigate the existence of significant spatial dependencies in the transform coefficients, and try to describe these dependencies using suitable long-tailed distributions [1]. A simple parametric model, augmented with a set of ”*hidden*” random variables that govern the model parameters such as local variance, models effectively these higher order dependencies. For instance, a hidden Markov model based on wavelet trees was successfully applied to signal [3] and image [4] denoising. Similar models have become widely used in speech processing [1]. On a unrelated front, interesting links between image compression and image denoising have been established. For instance, Natarajan [5] proposes the use of lossy compression for image denoising. The basic intuition behind the use of lossy compression for denoising is that a ”*signal typically has a correlated structure but white noise does not have structural redundancies*” [2]. The same intuition is also apparent in mathematically sound approaches such as minimum description length (MDL) and complexity regularized denoising [2]. Chang et al. [6] propose an adaptive model for image denoising via wavelet thresholding using context modeling of the *global* coefficients histogram. In this paper, we propose an approach to model neighboring coefficients of balanced multiwavelet transform. The model is inspired by the estimation-quantization (EQ) model that has been successfully applied in image coding [7]. Borrowing ideas from lossy image compression technology, denoising schemes based on the latter have been proposed [6]. These schemes exploit the intimate link that exists between lossy compression and denoising. Intuitively, images have structural correlations that an efficient image coder can exploit to reduce the existing statistical redundancy. However, white noise does not have any structural redundancy. Therefore, in the case of denoising, lossy compression allows for discriminating noisy regions from clear ones [6]. In this paper, we develop a local denoising solution where the image multiwavelet coefficients are modeled as zero-mean Gaussian random variables with high local correlation. The novelty of the proposed scheme lies in the use of balanced multiwavelet representations for image denoising. Unlike scalar wavelets [1], this class of representation allows for simultaneous orthogonality and filter symmetry.

II. Statistical Image Models for Denoising

Models of image statistics have been developed in early works of television engineering [1]. In these models, the the image statistics are assumed to be strict-sense stationary (i.e., spatially homogeneous). Also, the statistics are supposed to be invariant to changes in spatial scale. These assumptions lead to the widely used model: images can be characterized by a Gaussian random field with variance inversely proportional to the frequency. Therefore, under the Gaussianity assumption, the optimal estimator is linear. On the other hand, it became clear, using casual observation, that images can be highly

L. Ghouti is with the Department of Computer Science, King Fahd University of Petroleum and Minerals, Dhahran 31261, Saudi Arabia. Email: ghouti@ccse.kfupm.edu.sa

A. Bouridane is with the School of Computer Science, Queen's University of Belfast, BT7 1NN, UK. Email: A.Bouridane@qub.ac.uk

[1] We refer to classical wavelets as scalar wavelets.

0-7803-9134-9/05/$20.00 ©2005 IEEE

inhomogeneous due to the mixture of features such as edges and smooth regions. It is noted that large-amplitude transform coefficients are sparsely distributed throughout the image subbands.

A. Two-Step Variance-Adaptive Models

Most of image denoising techniques are based on Lee's two-step procedure [8]. In Lee's procedure, the local variance is first estimated from a neighborhood of pixels, then a linear least squares solution is derived using the estimated variance. This procedure can be cast into the class of empirical Bayes estimators [1] where the parameters of a local model are first estimated from the observed data, then a signal estimate is subsequently obtained using the estimated local parameter. Most of the variance-adaptive models published in the literature estimate the local variance from a set of neighboring subband coefficients, and then incorporate the estimated variance to denoise the observed image. On the other hand, GSM-based models yield excellent denoising results w.r.t to visual quality and in terms of mean squared error. It is worth noting that the class of two-step empirical Bayes approach is *suboptimal* [1], regardless of the optimality of the local variance estimator, because the estimation error associated with the first step is not incorporated in the second step. Li and Orchard [9] use a maximum a posteriori (MAP) estimator. Using a MAP estimator, but unlike [9], Mihçak et al. [2] define an exponential marginal prior, whereas Portilla et al. use a lognormal prior [10]. In [11], the set of multiplier variables has a global description modeled by a tree-structured Markov model. In this paper, we propose the use of balanced multiwavelet transforms with a *least squares optimal single-step* Bayesian estimator. Balanced multiwavelets, a class of multiresolution analysis representations based on multiple scaling and wavelet functions, achieve simultaneous symmetry and orthogonality. Specifically, these transforms can be characterized by time-varying filter banks [12]. However, it is well known that denoising schemes based on overcomplete representations perform better. In this case, a commonly followed solution to this problem is to eliminate the decimation operation [1].

1) Generalized Gaussian Model: In this work, we model image multiwavelet coefficients as a realization of a doubly stochastic process. In this case, the multiwavelet coefficients are assumed to be conditionally independent zero-mean Gaussian random variables, given their variances. These variances are modeled as identically distributed and highly correlated random variables. The multiwavelet coefficients are assumed to be "*locally*" independent and identically-distributed (i.i.d) variables. The two-step variance-adaptive denoising framework, considered in this paper, is defined by:

$$\mathbf{y} = \mathbf{x} + \mathbf{n} \tag{1}$$

where $\mathbf{x}$ and $\mathbf{y}$ represent the original and contaminated image pixels, respectively. The additive noise, $\mathbf{n}$, is additive white Gaussian (AWGN) with known variance σ_n^2. The multiwavelet equivalent of (1) yields quantities with similar statistical properties due to the orthonormality of the balanced multiwavelet transforms [12]. The multiwavelet coefficients of the contaminated image are given by $Y(k) = X(k) + w(k)$, where $w(k)$ is AWGN as explained before. The MMSE estimator for $X(k)$ is given by [2]:

$$\hat{X}(k) = \frac{\sigma_X^2(k)}{\sigma_X^2(k) + \sigma_n^2(k)} Y(k) \tag{2}$$

where $\sigma_X^2(k)$ represents the variance of the clean image coefficient. Because $\sigma_X^2(k)$ is not available, we approximate it using local maximum a posteriori (MAP) estimation procedure for each multiwavelet coefficient. The MAP procedure is performed using a local neighborhood and a prior model for $\sigma_X^2(k)$. Using the variance estimate, $\hat{\sigma}_X^2(k)$, the MMSE estimate is:

$$\hat{X}(k) = \frac{\hat{\sigma}_X^2(k)}{\hat{\sigma}_X^2(k) + \sigma_n^2(k)} Y(k) \tag{3}$$

2) Maximum Likelihood Estimation of $\hat{\sigma}_X^2(k)$: Using a square neighborhood of each multiwavelet coefficient, an estimate of $\hat{\sigma}_X^2(k)$ is derived assuming that the correlation between variances of neighboring coefficients is high. Therefore, the approximate Maximum Likelihood (ML) estimator is given by [2]:

$$\begin{aligned} \hat{\sigma}_X^2(k) &= \arg\max_{\sigma_X^2 \geq 0} \prod_{j \in \Omega(k)} P\left(Y(j)|\sigma_X^2\right) \\ &= \max\left(0, \frac{1}{M} \sum_{j \in \Omega(k)} Y^2(j) - \sigma_n^2\right) \end{aligned} \tag{4}$$

where $\Omega(k)$ is a square window centered at the multiwavelet coefficient $Y(k)$, $P\left(\cdot|\sigma_X^2\right)$ is the Gaussian distribution with zero-mean and variance $\sigma_X^2 + \sigma_w^2$, and M is the number of coefficients in the square window $\Omega(k)$. Using a prior marginal distribution prior $f_\sigma\left(\sigma_X^2\right)$ for each $\sigma_X^2(k)$, the approximate maximum *a posteriori* (MAP) estimator of $\sigma_X^2(k)$ is given by [2]:

$$\hat{\sigma}_X^2(k) = \arg\max_{\sigma_X^2 \geq 0} \left[\prod_{j \in \Omega(k)} P(Y(j)|\sigma_X^2)\right] f_\sigma(\sigma_X^2) \tag{5}$$

Using the exponential prior, $f_\sigma\left(\sigma_X^2\right) = \lambda e^{-\lambda \sigma_X^2}$, the MAP estimate of (5) becomes [2]:

$$\hat{\sigma}_X^2(k) = \max\left(0, \frac{M}{4\lambda}\left[-1 + \sqrt{1 + \frac{8\lambda}{M^2} \sum_{j \in \Omega(k)} Y^2(j)}\right] - \sigma_n^2\right) \tag{6}$$

III. Multiwavelets and Balanced Multiwavelets

Orthogonality is a desirable property for software/hardware implementation and symmetry provides comfort to image perception [12]. In the context of image processing applications, the following three properties are important: 1) orthogonality to ensure the decorrelation of subband coefficients, 2) symmetry (i.e., linear phase) to process finite length signals without redundancy and artifacts, and 3) finite-length filters for computational efficiency. However, most real scalar wavelet transforms fail to possess these properties simultaneously. To circumvent these limitations, multiwavelets have been proposed where orthogonality and symmetry are allowed to co-exist by relaxing the time-invariance constraint [12].

A. Multiwavelets

Multiwavelets may be considered as generalization of scalar wavelets. However, some important differences exist between these two types of multiresolution transforms. In particular, whereas wavelets have a single scaling $\phi(t)$ and wavelet function $\psi(t)$, multiwavelets may have two or more scaling and wavelet functions. In general, r scaling functions can be written using the vector notation $\Phi(t) = [\phi_1(t)\phi_2(t)\cdots\phi_r(t)]^T$, where $\Phi(t)$ is called the *multiscaling* function. In the same way, we can define the *multiwavelet* function using r wavelet functions as $\Psi(t) = [\psi_1(t)\psi_2(t)\cdots\psi_r(t)]^T$. The scalar case is represented by $r = 1$. Most of developed multiwavelet transforms use two scaling and wavelet functions, while r can take theoretically any value. Similar to scalar wavelets and for the case where $r = 2$, the multiscaling function satisfies the following two-scale equation:

$$\Phi(t) = \sqrt{2} \sum_{k=-\infty}^{\infty} H_k \Phi(2t-k), \tag{7}$$

$$\Psi(t) = \sqrt{2} \sum_{k=-\infty}^{\infty} G_k \Phi(2t-k), \tag{8}$$

However, it should be noted that $\{H_k\}$ and $\{G_k\}$ are 2×2 matrix filters defined as:

$$H_k = \begin{bmatrix} h_0(2k) & h_0(2k+1) \\ h_1(2k) & h_1(2k+1) \end{bmatrix}, \tag{9}$$

$$G_k = \begin{bmatrix} g_0(2k) & g_0(2k+1) \\ g_1(2k) & g_1(2k+1) \end{bmatrix} \tag{10}$$

where $\{h_k(n)\}$ and $\{g_k(n)\}$ are the scaling and wavelet filter sequences such that $\sum_n h_k^2(n) = 1$ and $\sum_n g_k^2(n) = 1$ for $k = 1, 2$. The matrix elements in the filters, given by (9) and (10), provide more degrees of freedom than scalar wavelets. However, the multi-channel nature of multiwavelets yields a subband structure that is different from that using scalar wavelets.

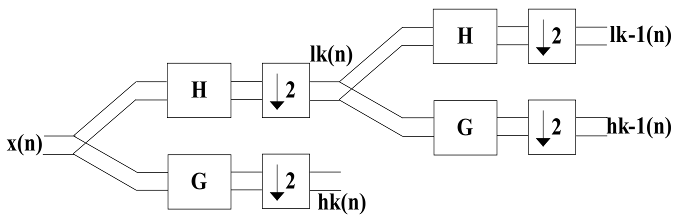

Fig. 1. Multiwavelet filter bank using one iteration.

Fig. 1 clearly shows that multiwavelets are defined for vector-valued signals (1D and 2D). Such *vectorizing* does not only introduce a fundamental asymmetry but it yields an approximation subband that does not represent a coarse approximation of the input image. The structure of the latter is different from that obtained using scalar wavelets. In the case of multiwavelets, the four sub-blocks of the approximation subband have very dissimilar spectral characteristics [12].

B. Balanced Multiwavelets

Lebrun and Vetterli [12] indicate that the balancing order of the multiwavelet is indicative of its energy compaction efficiency. However, a high balancing order alone does not ensure good image compression performance. For a scalar wavelet, the number of vanishing moments of its wavelet function $\int t^m \psi(t)dt = 0$ determines its vanishing order. For a scalar wavelet with vanishing order K, the highpass branch cancels a monomial of order less than K and the lowpass branch preserves it. For a multiwavelet transform, we have the similar notion of approximation order; a multiwavelet is said to have an approximation order of K if the vanishing moments of its wavelets, $\int t^m \psi_i(t)dt = 0$ for $0 \leq i \leq r-1$ and $0 \leq m \leq K-1$. An approximation order of K implies that the highpass branch cancels monomials of order less than K. But, in general, for multiwavelets, the preservation property does not automatically follow from the vanishing moments property. If the multifilter bank preserves the monomials at the lowpass branch output, the multiwavelet is said to be *balanced* [12]. The *balancing* order is p if the lowpass and highpass branches in the filter bank preserve and cancel, respectively all monomials of order less than p ($p \leq K$). Multiwavelets that do not satisfy the preservation/cancellation property are said to be *unbalanced*. For unbalanced multiwavelets, the input needs suitable *prefiltering* to compensate for the absence of the preservation/cancellation property, balancing obviates the need for input prefiltering; thus, they are computationally more efficient than the unbalanced multiwavelets. A time-varying representation of balanced multiwavelets is shown in Fig. 2.

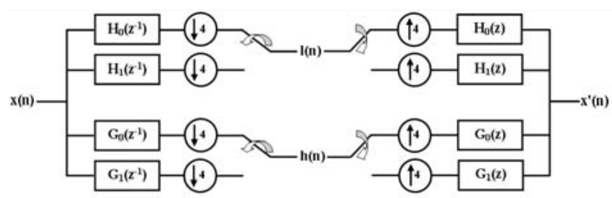

Fig. 2. Time-varying multiwavelet filter bank. Analysis stage (left). Synthesis stage (right).

The time-varying filter bank of Fig. 2 is defined by the following relations:

$$\begin{bmatrix} H_0(z) \\ H_1(z) \end{bmatrix} = H(z^2) \begin{bmatrix} 1 \\ z^{-1} \end{bmatrix} \tag{11}$$

$$\begin{bmatrix} G_0(z) \\ G_1(z) \end{bmatrix} = G(z^2) \begin{bmatrix} 1 \\ z^{-1} \end{bmatrix} \tag{12}$$

where $H_0(z)$ and $H_1(z)$ are the Z-transforms of the two lowpass branch filters h_0 and h_1. Similarly, $G_0(z)$ and $G_1(z)$

are the Z-transforms of the two highpass branch filters g_0 and g_1. In the time-varying filter bank implementation, the coefficients of the two lowpass (highpass) filters are simply interleaved at the output (see Fig. 2). Therefore, a separable 2D transform can now be defined in the usual way as the tensor product of two 1D transforms [12].

IV. Image Denoising Results

The denoising framework, considered in this paper, is defined by (1). The denoising procedure is summarized as follows: 1) apply 5-level balanced multiwavelet decomposition of the noisy image; 2) process all image subbands except the approximation subband; and 3) apply the inverse balanced multiwavelet to obtain an estimate of the denoised image. We assume that the noise $\mathbf{n}$ is an additive white Gaussian process of known variance. The estimation procedure is given by (2)-(4). We tested our algorithm on a number of images contaminated with synthetic Gaussian white noise at four different variances (10, 15, 20, and 25). Each noisy image is decomposed into five levels using BAT-1 family of balanced multiwavelets [12]. Different estimates of $\sigma_X^2(k)$ were obtained using centered square-shape windows of sizes 3×3, 5×5 and 7×7. Similar to [2], we set the parameter λ in (6) to the inverse of the standard deviation of the multiwavelet coefficients that are initially denoised using (4). In this paper, we compare our method to three different denoising schemes. The first one is the hard-thresholding of subband coefficients using a constant threshold for all subbands. The second method is MATLAB's denoising algorithm *wiener2* that implements a local variance-adaptive scheme in the pixel domain. While the last scheme is based on the two-step variance-adaptive model proposed by Mihçak et al. [2]. The peak signal-to-noise ratio (PSNR) results are shown in Table I. In this table, the proposed method, based on (6), is denoted *2VAR-BMWP-MAP*. We report only results for Lena and Barbara images. The results clearly indicate that the proposed 2VAR-BMWP-MAP scheme outperforms one of the best available published two-step denoising models [2]. Fig. 3(a) shows the original Barbara image. The same image, contaminated with white Gaussian noise of standard deviation $\sigma_n = 20$ is shown in Fig. 3(b). The denoised image using the 2VAR-BMWP-MAP denoising procedure is shown in Fig. 3(c). The results, shown in Table I and Fig. 3, apply a centered square-shape window of size 5×5 for the variance estimation. Finally, it should be noted that the proposed scheme has not been compared to the redundant multiresolution technique proposed in [1]. Such a comparison would require a formulation of balanced multiwavelet transform different from that used in this work.

Acknowledgment: The test images were provided by J. Portilla and I. Kozintsev. The authors would like to thank I. Kozintsev for providing the simulation code of [2]. L. Ghouti acknowledges the support of KFUPM University.

TABLE I

PSNR results in Decibels for several denoising methods

	Noise Standard Deviation σ_w			
	10	15	20	25
	Lena			
Donoho's hard thresholding [2]	30.34	28.52	27.24	26.34
MATLAB's *wiener2* [2]	32.98	30.44	28.52	26.95
Mihçak 5×5 LAWMAP [2]	34.31	32.36	31.01	29.98
2VAR-BMWP-MAP	**34.75**	**32.73**	**31.35**	**30.32**
	Barbara			
Donoho's hard thresholding [2]	27.29	25.01	23.65	22.83
MATLAB's *wiener2* [2]	31.35	28.58	26.67	25.19
Mihçak 5×5 LAWMAP [2]	32.57	30.19	28.59	27.42
2VAR-BMWP-MAP	**32.86**	**30.52**	**28.85**	**27.75**

(a) (b) (c)

Fig. 3. Denoising result on Barbara image. (a) Clean image. (b) Noisy image ($\sigma_n = 20$). (b) Denoised image using proposed 2VAR-BMWP-MAP scheme ($PSNR = 28.85\ dB$).

References

[1] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli, "Image denoising using scale mixtures of Gaussians in the wavelet domain," *IEEE Trans. on Image Processing*, vol. 12, no. 11, pp. 1338-1351, Nov. 2003.

[2] M. K. Mihçak, I. Kozintsev, K. Ramchandran, and P. Moulin, "Low-complexity image denoising based on statistical modeling of wavelet coefficients," *IEEE Signal Processing Letters*, vol. 6, no. 12, pp. 300-303, Dec. 1999.

[3] M. Crouse, R. Nowak, and R. Baraniuk, "Wavelet-based statistical signal processing using hidden Markov models," *IEEE Trans. Signal Processing*, vol. 42, no. 4, pp. 886902, Apr. 1998.

[4] J. Romberg, H. Choi, and R. Baraniuk, "Bayesian tree-structured image modeling using wavelet-domain hidden Markov models," in *Proc. SPIE Tech. Conf. Mathematical Modeling, Bayesian Estimation, and Inverse Problems*, Denver, CO, July 1999.

[5] B. Natarajan, "Filtering random noise from deterministic signals via data compression," *IEEE Trans. Signal Processing*, vol. 43, no. 11, pp. 25952605, Nov. 1995.

[6] S. G. Chang, B. Yu, and M. Vetterli, "Spatially adaptive wavelet thresholding with context modeling for image denoising," *IEEE Trans. Image Processing*, vol. 9, no. 9, pp. 15221531, Sept. 2000.

[7] 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.

[8] J. S. Lee, "Digital image enhancement and noise filtering by use of local statistics," *IEEE Pattern Anal. Machine Intell.*, vol. PAMI-2, pp. 165168, Mar. 1980.

[9] X. Li and M. T. Orchard, "Spatially adaptive image denoising under overcomplete expansion," in *Proc. IEEE Int. Conf. Image Processing*, Vancouver, BC, Canada, Sept. 2000.

[10] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli, "Adaptive Wiener denoising using a Gaussian scale mixture model in the wavelet domain," in *Proc. 8th IEEE Int. Conf. Image Processing*, pp. 37-40, Thessaloniki, Greece, Oct. 2001.

[11] 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. 89123, July 2001.

[12] J. Lebrun and M. Vetterli, "Balanced multiwavelets theory and design," *IEEE Trans. on Signal Processing*, vol. 46, no. 4, pp. 1119-1124, Apr. 1998.