

SPATIALLY ADAPTIVE STATISTICAL MODELING OF WAVELET IMAGE COEFFICIENTS AND ITS APPLICATION TO DENOISING

M. Kwanç Mihçak, Igor Kozintsev, Kannan Ramchandran

University of Illinois at Urbana-Champaign
Beckman Institute and ECE Department
405 N. Mathews Ave., Urbana, IL 61801
Email: {mihcak,igor,kannan}@ifp.uiuc.edu

ABSTRACT

This paper deals with the application to denoising of a very simple but effective “local” spatially adaptive statistical model for the wavelet image representation that was recently introduced successfully in a compression context [1]. Motivated by the intimate connection between compression and denoising [2, 3, 4], this paper explores the significant role of the underlying statistical wavelet image model. The model used here, a simplified version of the one in [1], is that of a *mixture process of independent* component fields having a zero-mean Gaussian distribution with unknown variances σ_s^2 that are slowly *spatially-varying* with the wavelet coefficient location s . We propose to use this model for image denoising by initially estimating the underlying variance field using a Maximum Likelihood (ML) rule and then applying the Minimum Mean Squared error (MMSE) estimation procedure. In the process of variance estimation, we assume that the variance field is “locally” smooth to allow its reliable estimation, and use an adaptive window-based estimation procedure to capture the effect of edges. Despite the simplicity of our method, our denoising results compare favorably with the best reported results in the recent denoising literature.

1. INTRODUCTION AND MOTIVATION

Accurate image modeling, whether done explicitly or implicitly, is a critical component to the effectiveness of almost all image processing tasks. For statistical-based approaches to image processing, the choice of an accurate statistical image model is critical. In [1], a simple yet effective statistical-based *spatially adaptive* wavelet image model was used in defining a compression algorithm (the so-called Estimation-Quantization (EQ) coder) whose performance ranks among the best in the image coding literature. In this work, we apply a simpler version of that model to the task of image denoising, motivated by the intimate connection between compression and denoising that has been established recently [2, 3, 4], which we will discuss shortly. Here, we use a simplified version of the statistical image model of [1] in formulating a simple method of denoising that attains state-of-the-art performance on standard test images. The key ingredient is the use of *spatial adaptivity* in a very simple statistical framework.

While we consider the specific case of additive white Gaussian noise (AWGN) in this work, extensions to more

general noise models are possible. In general, the denoising process involves the task of removing most of the artifacts due to noise while leaving the most important image components nearly undistorted. We choose the mean-squared-error (MSE) as the performance measure.

To motivate our approach, we discuss the relationship between image compression and image denoising. Historically, the problem of image compression has been the area of larger mass appeal and research activity. Partly as a result of this, the state-of-the-art in the image compression field seems to be the more mature of the two, especially with regard to the use of sophisticated and more realistic “real world” image models. On the other hand, the underlying relationship between the denoising and compression problems implies that the modeling maturity inherent in the successful state-of-the-art image coding algorithms may be effectively applied for the denoising problem as well. In fact, lossy data compression was proposed for denoising in [2]. The intuition behind this approach is that the “typically correlated signal is compressible but uncorrelated noise is not.” This principle, in a more rigorous form, is the basis for so-called complexity regularized denoising algorithms. Indeed, many of the recently proposed promising statistical denoising methods owe their improved performance to the more powerful image models that they derive inspired by state-of-the-art image compression algorithms. In the following, we describe some of the statistical image models in compression, starting from the most simple one, and we show how these models can be applied for image denoising.

A powerful class of image compression algorithms is based on the Discrete Wavelet transform (DWT) which effectively performs energy compaction for typical images by packing most of the image information into a few transform coefficients. The transform coefficients can be modeled as i.i.d. random variables with Generalized Gaussian distribution (first proposed in [5]) and this simple model can be used for designing the source coder. Similar models are used in several abovementioned complexity regularized denoising algorithms [6, 3]. These algorithms effectively perform global thresholding of wavelet coefficients, introduced in [7], by retaining only large coefficients and setting the rest to zero. Though wavelet thresholding methods have beautiful theoretical properties for certain classes of statistical models, they have not been very effective on “real-life” images primarily due to the mismatch in the modeling assumption.

Specifically, the lack of spatial adaptivity significantly reduces the performance of the discussed algorithms [6, 3, 7]. Our proposed model overcomes this disadvantage and, as a result, achieves a much better performance. It is interesting to note a similar evolution in the performance of wavelet image coders due to the spatially adaptive component in the subband image models.

That is, more sophisticated models for image compression recognize that there exist significant spatial dependencies in the transform coefficients and try to describe them using various data structures such as zero-trees introduced in [8]. The performance of the zero-tree based wavelet coders is currently among the best in image compression field. This inspired the proposal of similar methods in the image denoising application. In [4], a Hidden Markov model based on a wavelet tree was proposed to perform denoising. This model represents a compromise between an i.i.d. assumption and an overly general 2-D random process model by trying to capture most of the spatial dependencies while maintaining reasonable complexity. In [4], however, no real-life image results were presented. Though zero-tree based coders offer very good performance in image compression, attempts to translate this performance into simple algorithms for image denoising have not yet been very successful.

An alternative approach has been introduced recently, which offers excellent compression performance, while using a very simple and efficient local model [1]. In this work, we propose to modify the model of [1] for the purpose of image denoising and we demonstrate the promise of this approach. In [9] a similar spatially adaptive model for wavelet image coefficients was used to perform image denoising via wavelet thresholding. Within the framework of [9], each wavelet coefficient is modeled as a random variable with a Generalized Gaussian (GG) distribution having unknown parameters. Parameter estimation is carried out by context modeling. Our work, while being similar in approach (it employs the spatially adaptive model), differs from [9] in the statistical methods used for denoising - we perform MMSE estimation rather than coefficient thresholding. The main message of the paper is to introduce a simple stochastic model for image wavelet coefficients and to show its excellent performance in the denoising of real-life images.

2. PROBLEM FORMULATION AND SOLUTION

2.1. Stochastic model for wavelet coefficients

In our model, we assume that there exists a deterministic unknown spatially varying variance field. Given this field, the wavelet coefficients are modeled to be independent Gaussian random variables having zero mean and the variance given by the underlying field. We assume that the variance field is smoothly changing. Hence we approximate the wavelet coefficients as “locally i.i.d.”. Within the framework considered in this paper, the image pixels are corrupted by additive white Gaussian noise (AWGN) uncorrelated with the data as illustrated in Fig. 1.¹ The

¹Since in this work we use an orthogonal wavelet decomposition, AWGN in spatial domain preserves its statistical properties in the frequency domain.

denoising problem is to retrieve the original image coefficients as well as possible from the noisy observations. Here, we assume that we know the noise variance σ_n^2 . In Fig. 1, $X(k)$ represents the wavelet coefficients of a “clean” image corrupted by additive i.i.d. Gaussian noise samples $n(k)$ to produce the observed wavelet coefficients of a noisy image $Y(k)$. In Fig. 1, we explicitly show the image coefficients $X(k)$ as being obtained from the multiplication of the outcomes of an i.i.d. Gaussian source by samples $\sigma(k)$ from an unknown variance field. The model of the wavelet coefficients as being locally i.i.d. Generalized Gaussian was proposed and successfully used in [1] in the context of image compression. There, the variance estimate for the current coefficient was predicted from a quantized casual neighborhood, and the appropriate quantizer and entropy coder were matched to this variance estimate. Similarly, the algorithm proposed in this work is based on the underlying variance field being reasonably smooth to allow its reliable estimation from the local neighborhood of each wavelet coefficient. Specifically, we perform the ML estimation of the variance of each coefficient separately, using the observed noisy data in its neighborhood. This estimate is then used in the MMSE estimation (instead of the “true” variance) of the original wavelet coefficient $X(k)$. Both of these steps are outlined in the next subsections starting from the denoising part first and illustrated in Fig. 2.

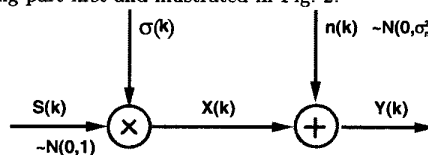


Figure 1: Model for noisy wavelet coefficients data. $X(k)$ represents wavelet coefficients of the “clean” image, each drawn independently from a Gaussian source with zero mean and variance $\sigma^2(k)$. $X(k)$ ’s are corrupted by AWGN samples $n(k)$ to produce the observed data $Y(k)$.

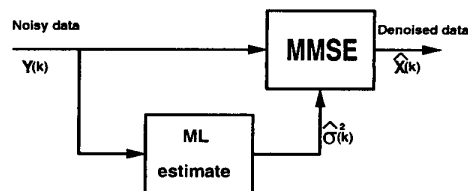


Figure 2: Block-diagram of the denoising algorithm. For each observed noisy coefficient $Y(k)$ we form the ML estimate $\hat{\sigma}^2(k)$ of the variance of the corresponding image wavelet coefficient $X(k)$ using its local neighborhood. Then, we find the MMSE estimate $\hat{X}(k)$ of the original image wavelet coefficient.

2.2. Denoising algorithm

Our proposed denoising algorithm is based on the Minimum Mean Squared error (MMSE) estimation procedure. Under the assumptions of independence and Gaussianity, the optimal (in the MSE sense) predictor for the clean data $X(k)$

is linear and is given by:

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

where $Y(k)$ is the observed data, $\sigma^2(k)$ is the variance of $X(k)$ and σ_n^2 is the variance of the AWGN samples. However, we cannot use the true signal variance $\sigma^2(k)$ in estimation since it is unknown. Instead, we propose to use the following form of the linear predictor:

$$\hat{X}(k) = \frac{\hat{\sigma}^2(k)}{\hat{\sigma}^2(k) + \sigma_n^2} Y(k) \quad (1)$$

where $\hat{\sigma}^2(k)$ is the estimated variance for the k -th data sample $X(k)$.

The performance of the proposed predictor (1) is dependent, to a high extent, on the performance of the estimator of the underlying variance field $\hat{\sigma}^2(k)$. Generally, this relation is unknown and complicated; in the interest of simplicity, we use the guiding principle 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 underlying variance field is the crux of the proposed denoising algorithm. For each data point (wavelet coefficient) an estimate of its variance is formed based on the local neighborhood of the data point. Under the smoothness assumption of the variance field, the transform coefficients around $X(k)$ can be used to estimate $\sigma(k)$. It is intuitively clear that the estimate of $\sigma(k)$ has to be dependent on the local data statistics. To formalize this idea, we consider the following setting. Suppose we are given a class of M different variance estimators $\Theta_m^2(k)$, only one of which will be used for each location k . To decide which estimator to use optimally, one needs to choose the estimator which minimizes the MSE $E[\hat{X}(k) - X(k)]^2$ where $\hat{X}(k)$ is obtained from (1) with $\hat{\sigma}^2(k) = \Theta_m^2(k)$. Unfortunately, this is an intractable problem, hence we propose to choose the estimator which minimizes the MSE in the estimation of the coefficient variance, i.e., we can select :

$$\hat{\sigma}(k) = \arg \min_{\Theta_m} E[(\sigma(k) - \Theta_m(k))^2], \quad (2)$$

which is easily shown to be equivalent to

$$\hat{\sigma}(k) = \arg \min_{\Theta_m} \{ (E[\Theta_m(k)] - \sigma(k))^2 + \text{Var}[\Theta_m(k)] \}. \quad (3)$$

The last equation is simply a bias-variance MSE decomposition.

In this work we consider only the specific class of window-type estimators of the variance of $X(k)$ where each data sample in the window around $X(k)$ contributes equivalently to the estimation. In general, to choose the estimator both the size and the shape of the neighborhood region should be determined. Obviously, given that the data samples are locally i.i.d., the bigger the size of the region to estimate $\sigma(k)$ is, the more reliable this estimate is. However, the locally i.i.d. assumption becomes inaccurate as the size of the neighborhood grows, and the quality of variance estimation may decrease, because of the possible inclusion of "wrong" data points into estimation. The outlined trade-off suggests the existence of an "optimal" neighborhood region for the

variance estimation for each data point. This region is data dependent and has to be determined. In the following, we present our approach to attack the problem of estimation of the variance field and the simulation results which validate the efficiency of our proposed methodology.

We consider M different local neighborhoods \mathcal{N}_{k_m} in order to estimate the variance $\sigma^2(k)$ for the transform coefficient $X(k)$, where $m \in \{1, 2, \dots, M\}$. We model the statistical properties of the transform coefficients within the set \mathcal{N}_{k_m} as being stationary i.i.d. Thus, the maximum-likelihood variance estimator for $X(k)$ using the neighborhood \mathcal{N}_{k_m} is:

$$\Theta_m^2(k) = \max \left(0, \sum_{l \in \mathcal{N}_{k_m}} Y^2(l) - \sigma_n^2 \right). \quad (4)$$

In our denoising algorithm, square-shaped windows of different sizes around the data has been used to estimate $\sigma^2(k)$ for simplicity. Due to a combination of the intractability of the minimization in (3) because of the presence of the first term as well as the dominance of the second term (which we verify in Table 1), we propose to modify the criterion (3) to choose the local estimator by considering only the variance term in the minimization. We found experimentally that this term dominates the estimator error therefore our criterion in simulations was:

$$\hat{\sigma}^*(k) = \arg \min_{\Theta_m} \text{Var}[\Theta_m(k)]. \quad (5)$$

In order to estimate the variance of the estimator, we use a simple technique, called the *Bootstrap Method*.

The Bootstrap is a practical technique for validating the accuracy of a parameter estimator and obtaining information about the distribution of the estimator. While other statistical methods having better asymptotical properties exist, they usually require a large number of samples for an efficient performance, which is clearly not the case in our situation. On the other hand, the Bootstrap provides a simple and efficient alternative under the constraint of a limited sample size.

To understand how the Bootstrap works in our case, suppose the estimator $\Theta_m(k)$ is obtained using the samples in the set \mathcal{N}_{k_m} . Let the corresponding data size be $|\mathcal{N}_{k_m}|$. The first step is the "Resampling" step, where a random sample of size $|\mathcal{N}_{k_m}|$ is drawn from the set \mathcal{N}_{k_m} with replacement. The second step is the calculation of the bootstrap estimate using this random sample (i.e., simply using (4) on this random sample). These two steps are repeated a large number of times. The outcomes of the second step form an approximation to the statistical distribution of the estimator $\Theta_m(k)$ from which the standard deviation of the estimator can be approximated. For more information on Bootstrap, the reader is referred to [10].

3. RESULTS AND DISCUSSION

We tested our algorithm on a number of images, here, we only report the results for *lena* and *barbara*. We generated i.i.d. Gaussian noise at four different values of the noise variance, σ_n^2 . We used orthogonal wavelet transform with five levels of decomposition and Daubechies' length-8 wavelet. Square-shaped windows of sizes 3×3 , 5×5 , 7×7 and 9×9 around the data point were employed to

find different estimates for $\sigma^2(k)$. For each transform coefficient, one of these estimators is favored as a result of our algorithm. A typical estimator assignment map is shown in Figure 3. Note that around edges, small size windows are selected because of the fact that the variations in the image data are considerable in these regions. On the other hand, in smoother background regions, windows with larger sizes are preferred. Five different methods have been com-



Figure 3: The estimator assignment map for all of the subbands of the *lena* image when $\sigma_n = 15$. Note how smaller window sizes for predictors are favored around edges, while larger window sizes for predictors are preferred in smooth background regions.

pared, and the PSNR results are shown in Table (1). The first method is the hard-thresholding of wavelet coefficients using a constant threshold for all subbands, calculated according to [7]. The second method is the image denoising algorithm of MATLAB, which is invoked by *wiener2*. The third method is based on a framework of spatially adaptive wavelet thresholding, which is a recent work [9]. We included only the results from [9] which were obtained by using an orthogonal wavelet transform since this is equivalent to our setup. However, it has been reported in [9] that improvements of 1 – 1.5 dB can be obtained by using the overcomplete expansion. We plan to extend our method to the non-subsampled wavelet transform domain in our future research.

For each image we also presented the results obtained from “optimal averaged estimator assignment maps”. These maps represent the empirical upper bound on PSNR performance for the class of estimators that we use. Note that the performance of our method is approaching the best performance possible for this class of estimators, which further justifies our estimator selection criterion (5). In this work, a very limited class of estimators was used. We believe that if a richer class of estimators is allowed, the performance would improve substantially. This is the subject of our ongoing research.

4. REFERENCES

[1] S. LoPresto, K. Ramchandran, and M. T. Orchard, “Image coding based on mixture modeling of wavelet

Table 1: PSNR results for several denoising methods

PSNR(dB)/ σ_n	10	15	20	25
<i>LENA</i>				
Donoho’s hard thresholding [7]	30.34	28.52	27.24	26.34
Spatial local Wiener (MATLAB)	32.54	30.44	28.52	26.95
SAWT [9]	—	31.83	30.49	29.50
Our method	34.39	32.38	30.96	29.84
Empirical upper bound	34.58	32.59	31.17	30.06
<i>BARBARA</i>				
Donoho’s hard thresholding [7]	27.29	25.01	23.65	22.83
Spatial local Wiener (MATLAB)	31.35	28.58	26.67	25.19
SAWT [9]	—	29.19	27.65	26.52
Our method	32.67	30.26	28.65	27.43
Empirical upper bound	32.84	30.46	28.86	27.65

coefficients and a fast estimation-quantization framework,” *Proc. IEEE Data Compression Conf.*, 1997.

- [2] B. Natarajan, “Filtering random noise from deterministic signals via data compression,” *IEEE Trans. on Signal Processing*, vol. 43, pp. 2595–2605, Nov. 1995.
- [3] J. Liu and P. Moulin, “Complexity-regularized image denoising,” in *Proc. IEEE Int. Conf. Image Proc.*, pp. 370–373, 1997.
- [4] M. Crouse, R. Nowak, and R. Baraniuk, “Wavelet-based statistical signal processing using hidden markov models,” *IEEE Trans. on Signal Processing*, vol. 42, april 1998.
- [5] S. Mallat, “A theory for multiresolution signal decomposition: the wavelet representation,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, pp. 674–693, July 1989.
- [6] 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.
- [7] D. L. Donoho and I. M. Johnstone, “Ideal spatial adaptation by wavelet shrinkage,” *Biometrika*, pp. 425–455, 1994.
- [8] J. M. Shapiro, “Embedded image coding using zerotrees of wavelet coefficients,” *IEEE Trans. Signal Process.*, vol. 41, pp. 3445–3462, Dec. 1993.
- [9] 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.
- [10] B. Efron and B. Tibshirani, *An Introduction to the Bootstrap*. Chapman and Hall, 1993.