SPATIALLY ADAPTIVE WAVELET THRESHOLDING WITH CONTEXT MODELING FOR IMAGE DENOISING

S. Grace Chang[1] *Bin Yu*[2] *Martin Vetterli*[1,3]

[1]Department of Electrical Engineering and Computer Sciences
University of California, Berkeley, CA 94720, USA
[2]Department of Statistics
University of California, Berkeley, CA 94720, USA
[3]Département d'Electricité
Ecole Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland
grchang@eecs.berkeley.edu, binyu@stat.berkeley.edu, vetterli@de.epfl.ch

ABSTRACT

The method of wavelet thresholding for removing noise, or denoising, has been researched extensively due to its effectiveness and simplicity. Much of the work has been concentrated on finding the best uniform threshold or best basis. However, not much has been done to make this method adaptive to spatially changing statistics which is typical of a large class of images. This work proposes a spatially adaptive wavelet thresholding method based on context modeling, a common technique used in image compression to adapt the coder to the non-stationarity of images. We model each coefficient as a random variable with the Generalized Gaussian prior with unknown parameters. Context modeling is used to estimate the parameters for each coefficient, which are then used to adapt the thresholding strategy. Experimental results show that spatially adaptive wavelet thresholding yields significantly superior image quality and lower MSE than optimal uniform thresholding.

1. INTRODUCTION

In this paper we address the classical problem of removing additive noise from a corrupted image, or *denoising*. In recent years there has been a plethora of work on using *wavelet thresholding* [3] for denoising, in both the signal processing and statistics community, due to its effectiveness and simplicity. In its most basic form, this technique denoises in the orthogonal wavelet domain, where each coefficient is *thresholded* by comparing against a threshold; if the coefficient is smaller than the threshold, it is set to zero, otherwise it is kept or modified. The intuition is that because the wavelet transform is good at energy compaction, small coefficients are more likely due to noise, and large coefficients due to important signal features (such as edges). The threshold thus acts as an oracle deciding whether or not to keep the coefficients. Most of the literature thus far has concentrated on developing threshold selection methods, with the threshold being uniform or at best one threshold for each subband. Very little has been done on developing thresholds that are adaptive to different spatial characteristics. Other works investigate the choice of wavelet coefficient expansion for the thresholding framework. One particularly interesting result is that thresholding in a shift-invariant expansion (dubbed *translation-invariant (TI) denoising* by Coifman and Donoho [2]) eliminates some of the unpleasant artifacts introduced by modifying the coefficients of the orthogonal wavelet expansion. In this paper, we use the wisdom that thresholding in a shift-invariant, overcomplete representation outperforms the orthogonal basis, and also investigate an issue that has not been explored, namely, the spatial adaptivity of the threshold value.

A spatially adaptive thresholding strategy is needed because sometimes a uniform threshold is not good enough. The essence of a threshold is that it should be large enough to kill the noise, but small enough to keep the signal features. However, when the noise coefficients happen to be larger than the signal coefficients, it may not possible to accomplish both goals with just one threshold. Thus, if we can extract additional information from the image to distinguish between the important and noisy coefficients, then adaptive thresholds can be used to reap both the benefits of keeping the important signal features while removing most of the noise.

Most natural images have non-stationary properties, since they typically consist of regions of smoothness and sharp transitions. These regions of varying characteristics can be well differentiated in the wavelet domain, as can be seen in the wavelet decomposition of the *lena* image in Figure 1. One observes areas of high and low energy (or large and small coefficient magnitude), represented by white and black pixels, respectively. Areas of high energy correspond to signal features of sharp variation such as edges and textures; areas of low energy correspond to smooth regions. When noise is added, it tends to increase the magnitude of the wavelet coefficient on average. Specifically, in smooth regions, one expects the coefficients to be dominated by noise, thus most of these coefficients should be removed, especially since noise is highly visible here. In regions of sharp transition, the coefficients have a lot of energy due to the signal, and some due to noise (which is not as visible in these regions), thus they should be kept, or at most modified only a little, to ensure that most of the signal details are retained. Thus, the idea is to distinguish between the low and high energy regions, and modify the coefficients using a spatially adaptive thresholding strategy.

It has long been accepted in the subband coding community that for a large class of images, the coefficients in each subband form a distribution well described by the Generalized Gaussian prior [7]. The classification-based compression method in [8] found that these coefficients can be further clustered into several subgroups, each described by this distribution but of different parameters. The clustering of the coefficients is based on context-modeling, a popular method used in compression for differentiating pixels of varied characteristics. Thus, context-modeling allows us to model each coefficient as a Generalized Gaussian random variable with varying parameters. Now, given that we can estimate the parameters for each coefficient, the next step is to use them to calculate the threshold. In [1], we found that when the signal coefficients are modeled as Generalized

0-8186-8821-1/98 $10.00 © 1998 IEEE

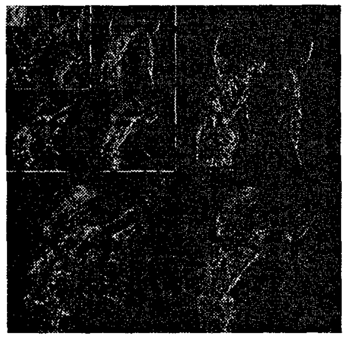

Figure 1. Four level wavelet decomposition of *lena*. White pixels indicate large magnitude coefficients, and black signifies small magnitude.

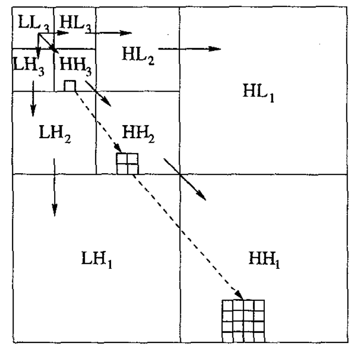

Figure 2. Subbands of the orthogonal discrete wavelet transform(DWT), also showing the parent-child relationship.

Gaussian random variables and the noise as Gaussian, the threshold $\tilde{T} = \sigma^2/\sigma_x$ is a good approximation to the optimal threshold which minimizes the mean squared error of the thresholding estimator, where σ^2 is the noise power, and σ_x is the standard deviation of the signal. The simplicity of this threshold makes it easy to achieve spatial adaptivity – one only needs to quantify the local characteristic in σ_x to make the threshold $\tilde{T}$ adaptive on a pixel-by-pixel manner.

Our proposed algorithm is based on using adaptive thresholding in the overcomplete wavelet expansion. This method outperforms both using only adaptive thresholding in the orthogonal expansion or using only uniform thresholding in the overcomplete expansion like the TI denoising. That is, by combining both features, we achieve results which are significantly more superior than either method alone. The organization of this paper is as follows. In Section 2, we introduce the threshold selection method when there is only one class of Generalized Gaussian distributed random variable corrupted by additive Gaussian noise. Because this threshold selection is based on *iid* noise assumption, the discussion will first set in the orthogonal wavelet transform. Then context modeling is introduced to allow each coefficient be modeled as random variables of different parameters, and the parameters are used to make the threshold spatially adaptive. Finally, we discuss how to extend this adaptive method in the orthogonal expansion to the overcomplete expansion. In Section 3, we will compare the spatially adaptive results with those from the best uniform thresholding strategy (in the mean squared error sense), in both the orthogonal and overcomplete expansion, and show that the combination of using spatially adaptive thresholding and overcomplete expansion yields superior results in both the visual quality and the mean squared error.

2. ADAPTIVE ALGORITHM

The adaptive algorithm will be developed in the following manner. First, we introduce the concept of modeling the orthogonal wavelet coefficients by the Generalized Gaussian prior (with unknown parameters), and develop a threshold selection method when the coefficients are corrupted by Gaussian noise. Then context modeling is used to model each coefficient as a random variable with different unknown parameters, allowing essentially an infinite mixture of distributions. The threshold for each coefficient is adjusted according to the estimated parameters for that coefficient. Lastly, since the aforementioned algorithm is developed in the orthogonal expansion where the coefficients are uncorrelated, we will need to take care in extending it to the overcomplete expansion, where coefficients are correlated.

2.1. Coefficient Modeling and Threshold Selection

Let the corrupted image model be

$$y[i,j] = f[i,j] + \varepsilon[i,j], i,j = 1,\ldots,N,$$

where $f[i,j]$ is the original image, and $\varepsilon[i,j]$ are *iid* $N(0,\sigma^2)$ and independent of $f[i,j]$. The goal is to recover $f[i,j]$ from the noisy observations $y[i,j]$.

To accomplish wavelet thresholding for denoising, the observations $\{y[i,j]\}$ are first transformed into the wavelet domain. The necessary notations for the wavelet transform will be introduced here, and the readers are referred to references such as [5, 6] for more details. The 2D discrete orthogonal wavelet transform (DWT) can be implemented as a critically sampled octave-band filter bank, where separable filtering is used. It is often convenient to cluster these coefficients into groups or *subbands* of different scales and orientations as in Figure 2, where, for example, the label HL_1 refers to those coefficients at the first scale of decomposition which are the output of the highpass filter in the horizontal direction and the lowpass filter in the vertical direction. Let $Y^{(s,o)}[i,j], i,j = 1,\ldots,N/2^s$, denote the wavelet coefficients of $y[i,j]$ at a particular scale s and orientation o, where $s = 1,2,\ldots,J$ and $o \in \{HL, LH, HH, LL\}$.

It has been observed that for a large class of images, the coefficients from each subband (except LL) form a symmetric distribution that is sharply peaked at zero, well described by the Generalized Gaussian distribution [7], $GG_{\alpha,\beta}(x) = C(\alpha,\beta)\, e^{-(\alpha|x|)^\beta}$, where $C(\alpha,\beta) = \frac{\alpha\beta}{2\Gamma(\frac{1}{\beta})}$ and $\Gamma(t) = \int_0^\infty e^{-u}u^{t-1}du$ is the gamma function. Let $X^{(s,o)}[i,j]$ and $V^{(s,o)}[i,j]$ denote the wavelet coefficients of the original signal $f[i,j]$ and the noise $\varepsilon[i,j]$, respectively. For each subband, the signal coefficients $X^{(s,o)}[i,j]$ are modeled as independent samples of distribution, $p_X(x) = GG_{\alpha,\beta}(x)$, and the noise as independent samples of the Gaussian distribution, $p_V(v) = \phi(v,\sigma^2) = 1/\sqrt{2\pi\sigma^2}\exp -\frac{v^2}{2\sigma^2}$. If we restrict the estimator to be a *soft-threshold* estimator of the form $\hat{X}^{(s,o)}[i,j] = \eta_T(Y^{(s,o)}[i,j])$, where $\eta_T(x) = \mathrm{sgn}(x)\cdot\max(|x|-T, 0)$, then the optimal threshold T^* is defined to be the argument which minimizes the expected squared error,

$$T^* = \arg\min_T E_{Y|X,X}(\eta_T(Y) - X)^2 \qquad (1)$$

where $Y|X \sim \phi(y - x, \sigma^2)$ and $X \sim GG_{\alpha,\beta}(x)$. In [1], we found that T^* can be well approximated by $\tilde{T} = \sigma^2/\sigma_x$, where σ_x is the standard deviation of X. Thus, by estimating the standard deviation of the signal coefficients in each subband, we have a uniform threshold that is adaptive to each subband characteristic. Note that it is not necessary to explicitly estimate the parameters α, β since the standard deviation suffices for our purpose.

The threshold $\tilde{T} = \sigma^2/\sigma_x$ is not only nearly optimal but also has an intuitive appeal. For such a choice, the normalized threshold $\tilde{T}/\sigma$ is inversely proportional to σ_x, the standard deviation of X, and proportional to σ, the noise standard deviation. When $\sigma/\sigma_x \ll 1$, the signal is much stronger than the noise, thus $\tilde{T}/\sigma$ is chosen to be small in order to preserve most of the signal and remove some of the noise; vice versa, when $\sigma/\sigma_x \gg 1$, the noise dominates and the normalized threshold is chosen to be large to remove the noise which has overwhelmed the signal. Thus, this threshold choice adapts to both the signal and noise characteristics reflected in the parameters σ and σ_x.

2.2. Context Modeling for Spatial Adaptivity

To make the threshold $\tilde{T}$ spatially adaptive, we need to develop a method to estimate σ_x locally, not just at the subband level. To do this, we adopt the *context modeling* idea used frequently in image compression for adapting the coder to changing image characteristics. That is, the statistical model for a given coefficient is conditioned on a function of its neighbors. In the wavelet-based compression scheme in [8], context modeling was used to further categorize coefficients into several classes of varied activity levels within each subband, that is, classes of Generalized Gaussian distribution with different parameters α, β. The distribution parameters are estimated from the coefficients for each class, which are then used to adapt the coder. Since the description of each class and the distribution parameters needs to be sent as overhead, only four classes were used in [8]. For the denoising problem, there is no need to conserve bits, thus it is not necessary to explicitly classify the pixels, and parameters can be estimated for each coefficient (rather than for each class), resulting in virtually an infinite mixture of distributions.

Consider one particular subband with M^2 coefficients, and a particular pixel $Y^{(s,o)}[i,j]$ at location index $[i,j]$. To simplify notation, we drop the superscript (s,o), and use it only when necessary for clarity. Each coefficient $Y[i,j]$ is modeled as a random variable whose variance can be estimated as follows. Consider a neighborhood of $Y[i,j]$, whose p elements are placed in a $p \times 1$ vector u_{ij}. One possible choice is the eight nearest neighbors of $Y[i,j]$ in the same subband, plus its parent coefficient $Y^{(s+1,o)}[\lceil i/2 \rceil, \lceil j/2 \rceil]$ (see Figure 2 for the definition of parent-child relationship). To characterize the activity level of the current pixel, we calculate a weighted average of the absolute value of the neighbors, $Z[i,j] = w^T\, |u_{ij}|$. The weights w is found by using the least squares estimate, that is,

$$w^* = \arg\min_{w} \sum_{i,j} (Y[i,j] - w^T\, |u_{ij}|)^2 = (|U^T|\, |U|)^{-1} |U^T|\, Y$$

where U is a $M^2 \times p$ matrix with each row being u_{ij}^T, for all i,j, and Y is the $M^2 \times 1$ vector containing all coefficients $Y[i,j]$.

The variance of coefficient $Y[i,j]$ is estimated from other coefficients whose context lie in the interval $[Z[i,j] - \Delta_1, Z[i,j] + \Delta_2]$. To develop an intuition for this, it is helpful to examine Figure 3, which plots the pairs $\{Z[i,j], Y[i,j]\}, i,j = 1,\ldots,M$. The points are clustered within a cone shape centered at origin. Taking an interval of small valued $Z[i,j]$, the associated coefficients $\{Y[i,j]\}$ have a small spread; on the other hand, an interval of large valued $Z[i,j]$ has corresponding $\{Y[i,j]\}$ with a larger spread (the intervals are of different widths to capture the same number of points). This suggests that the context provides a good indication of local variability. Thus, for a given coefficient $Y[i_0, j_0]$, we place an interval around $Z[i_0, j_0]$, and the variance of $Y[i_0, j_0]$ is estimated from the points $Y[i,j]$ whose context falls within this window. In particular, we take L closest points above $Z[i_0, j_0]$ and L closest points below, resulting in a total of $2L + 1$ points, where we choose $L = \max(50, M^2/10)$ to ensure that enough points are used to estimate the variance. Note that this is a moving window rather than the fixed classes in [8], and thus allows a continuous range of estimate values. Let $\mathcal{B}_{i_0 j_0}$ denote the set of points $\{Y[i,j]\}$ whose context falls in the moving window. The estimate of the variance $\sigma_x^2[i_0, j_0]$ is then

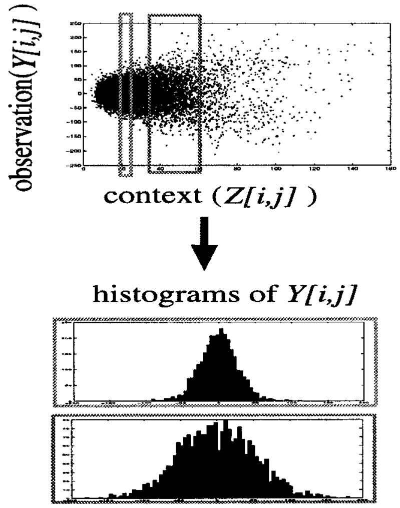

Figure 3. A sample plot of $\{Z[i,j], Y[i,j]\}$, where $Y[i,j]$ is the noisy wavelet coefficient, and $Z[i,j]$ is its context. A collection of $Y[i,j]$ with small values of $Z[i,j]$ have a smaller spread than those with large values of $Z[i,j]$, suggesting that context modeling provides a good variability estimate of $Y[i,j]$.

$$\hat{\sigma}_x^2[i_0, j_0] = \max\left(\frac{1}{2L+1} \sum_{Y[k,\ell] \in \mathcal{B}_{i_0 j_0}} Y[k,\ell]^2 - \sigma^2, 0 \right).$$

The term σ^2 needs to be subtracted because $\{Y[i,j]\}$ are the noisy observations, and the noise is independent of the signal, with variance σ^2. The threshold at location $[i_0, j_0]$ is then

$$\tilde{T}_{i_0 j_0} = \frac{\sigma^2}{\hat{\sigma}_x[i,j]}.$$

Calculating the threshold $\tilde{T}_{ij}$ for every location $[i,j]$ yields a spatially adaptive thresholding strategy. In the implementation, the context $\{Z[i,j]\}$ are first sorted, and a moving window is placed over them, so the set $\mathcal{B}_{ij}$ and the variance estimate $\hat{\sigma}_x^2[i,j]$ can be updated efficiently.

There are several noteworthy remarks to be made about our proposed approach. One may ask why the local variance is not estimated from, say, a local window, but rather from an indirect way of grouping the coefficients first via its context. Estimating from a local neighborhood is simple, and, as demonstrated by the good performance of the image coder in [4], it yields an estimate good enough for adapting the coder. However, our experience with noisy images show that such an estimate yields considerably more unreliable variance estimates and also blotchy denoised image. This is because the estimate is highly sensitive to the window size we choose: a small window contains few points and thus yields unreliable estimates; a large window adapts slowly to different characteristics. The context-based grouping allows one to congregate those coefficients with similar context though not necessarily spatially adjacent. It also allows a large number of coefficients to be used in the variance estimation, thus yielding a more reliable estimate. Via some simulations, we find that the neighborhood choice $\mathcal{B}_{ij}$ and the weight w used in the context calculation is not very sensitive, as a simple equally weighted average of the eight nearest neighbors yield approximately the same result.

Up to now we have not discussed how to estimate the noise variance σ^2. In some practical cases, it is possible to measure σ^2 based on information other than the corrupted observation. If this is not the case, we estimate it by using the robust median estimator in the highest subband of the wavelet transform, $\hat{\sigma} = \text{Median}(|Y[i,j]|)/.6745$, $Y[i,j] \in$ subband HH_1, also used in [3].

2.3. Thresholding in Overcomplete Expansion

Thresholding in the orthogonal wavelet domain produces significantly noticeable artifacts such as Gibbs-like ringing and blips. To ameliorate this unpleasant phenomenon, Coifman and Donoho [2] proposed the *translation-invariant (TI) denoising*. Let $\textit{Shift}_{k,\ell}[\cdot]$ denote the operation of circularly shifting the input by k indices in the vertical direction and ℓ indices in the horizontal, and let $\textit{Unshift}_{k,\ell}[\cdot]$ be a similar operation but in the opposite direction. Also, let $\textit{Denoise}[\cdot,T]$ denote the operation of taking the DWT of an input signal, threshold it with a chosen uniform threshold T, then transform it back to the space domain. Then the TI denoising yields an output which is the average of the thresholded copies over all possible shifts: $\hat{f} = \frac{1}{N^2}\sum_{k,\ell=0}^{N-1} \textit{Unshift}_{k,\ell}[\textit{Denoise}[\textit{Shift}_{k,\ell}[y],T]]$. The rationale is that since the orthogonal wavelet transform is a time-varying transform and thresholding the coefficients produces ringing-like phenomena, thresholding a shifted input would produce ringing at different locations, and averaging over all different shifts would yield an output with more attenuated artifacts than a single copy alone. TI denoising can be shown to be equivalent to thresholding in the overcomplete representation implemented by the *non-subsampled filter bank* as will be described below, up to some scaling in the thresholds. It has been shown to remove some of the ringing artifacts, because denoising in the redundant expansion can be interpreted as an additional averaging. Thus we proceed to extend our spatial adaptive algorithm to this redundant expansion.

The adaptive algorithm in the orthogonal basis described above can easily be extended to the overcomplete basis. Now consider the same orthogonal filters but used in a filter bank without downsampler (see [6] for more detail on non-subsampled filter banks). The filters are renormalized by $1/\sqrt{2}$ so that coefficient energy stays bounded. This decomposition is a redundant representation, and there are correlations between the decomposition coefficients. Specifically, at the first level of decomposition, the odd and even coefficients are correlated. Thus, we can separate the coefficients into four sets of uncorrelated coefficients, namely, $\{Y[2i,2j]\}, \{Y[2i,2j+1]\}, \{Y[2i+1,2j]\}$ and $\{Y[2i+1,2j+1]\}$. For the s-th level decomposition, the coefficients can be separated into 2^{2s} sets, each containing uncorrelated coefficients, and they are $\{Y[2^s i+k_1, 2^s j+k_2]\}_{i,j}, k_1,k_2 = 0,1,\ldots,2^s-1$. Since each set contains uncorrelated coefficients, the noise are also *iid* within each set as well, and thus the adaptive algorithm can be used for each set of coefficients. This approach let us circumvent the issue of denoising correlated coefficients with colored noise, which is not an easy task. There are several other minor details in this implementation. Firstly, one needs to alter the noise power σ^2 at each decomposition scale to $\sigma^2/4^s$ due to the renormalization of the filters. Secondly, the definition of the parent coefficient used in the neighborhood of the context is slightly changed: the parent of a coefficient in scale s is simply the coefficient at the same spatial location in scale $s+1$.

Table 1. Comparing the MSE of the spatially adaptive algorithm with optimal subband uniform threshold in the DWT and overcomplete expansion for various test images and σ.

MSE/σ	12.5	15	20	22.5	25
	barbara				
AdaptDWT	61.4	78.3	111.6	127.5	144.8
OrcUnifDWT	62.2	80.7	117.3	136.8	155.0
AdaptNS	**43.5**	**56.0**	**83.1**	**97.5**	**112.2**
OrcUnifNS	51.2	66.3	96.7	112.0	128.2
	lena				
AdaptDWT	36.1	42.7	58.1	66.5	72.9
OrcUnifDWT	36.1	43.7	58.8	67.4	73.7
AdaptNS	**27.5**	**32.7**	**44.1**	**51.1**	**56.5**
OrcUnifNS	29.8	35.9	48.7	55.7	61.2

3. EXPERIMENTAL RESULTS

We use the images *barbara* and *lena* as test images. *iid* Gaussian noise at different levels of σ^2 are generated using *randn* in MATLAB. For the orthogonal wavelet transform, four levels of decomposition are used, and the wavelet employed is Daubechies' symmlet with 8 vanishing moments. There are four methods that we compare, and the MSE results are shown in Table 1. The *AdaptDWT* method refers to the proposed adaptive thresholding using the orthogonal transform DWT, and *AdaptNS* refers to adaptive thresholding using the non-subsampled wavelet transform. These two are compared against the best uniform thresholding techniques (in the MSE sense) when the original uncorrupted image is assumed to be known. For thresholding with DWT, in each subband, we find the *oracle* threshold T_{orc} as

$$T_{orc} = \arg\min_T \|\eta_T(Y[i,j]) - X[i,j]\|^2$$

where $Y[i,j]$ and $X[i,j]$ are the wavelet coefficients of the noisy observation y and original image f, respectively. This method is labeled *OrcUnifDWT* in Table 1. Similarly, this is extended to the non-subsampled wavelet transform, where a different threshold is found for each set of uncorrelated coefficients within each subband (thus 2^{2s} thresholds for a subband at scale s). This method is labeled *OrcUnifNS*. Figure 4 shows a magnified region in the *barbara* image for $\sigma = 25$. The *AdaptNS* method outperforms all the other methods in both visual quality and MSE performance. It yields significantly less ringing artifacts and blotchiness than the methods using DWT. The *OrcUnifNS* method using uniform thresholds in the non-subsampled framework still shows significant noise in the smooth background. Thus, it is both the spatial adaptive thresholds and the overcomplete representation that contribute to the superior quality of the *AdaptNS* method. The adaptive methods denoise better especially in the flat regions, where the uniform methods yields images with much noise and "blips". Note that although the MSEs for the *lena* image is similar between the adaptive and uniform oracle methods, the visual quality in the adaptive method is far superior as it produces a denoised image that is smooth in the flat regions and has less artifacts around the edges as well.

4. CONCLUSION

We have proposed a simple and effective spatially and scalewise adaptive method for denoising via wavelet thresholding in the overcomplete expansion. The issue of spatially adapting the threshold values has not been addressed in the literature. As we have shown in this paper, adapting the threshold values to local signal energy allows us to keep much of the edge and texture details, while eliminating most of the noise in smooth regions. The results shows substantial improvement over the best uniform thresholding, both in visual quality and mean squared error.

REFERENCES

[1] S.G. Chang, B. Yu, and M. Vetterli, "Image Denoising via Lossy Compression and Wavelet Thresholding," *Proc. IEEE Int. Conf. Image Processing*, Vol.1, pp. 604-607, Oct. 1997.

[2] R.R. Coifman and D.L. Donoho, "Translation-invariant de-noising," Wavelets and Statistics, A. Antoniadis and G. Oppenheim eds., Springer-Verlag Lecture Notes, 1995.

[3] D.L. Donoho and I.M. Johnstone, "Ideal spatial adaptation via wavelet shrinkage," *Biometrika*, vol 81, pp. 425-455, 1994.

[4] S. LoPresto, K. Ramchandran, and M. Orchard, "Image coding based on mixture modeling of wavelet coefficients and a fast estimation-quantization framework," *Proc. Data Compression Conference*, Snowbird, Utah, March 1997.

[5] S. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Pat. Anal. Mach. Intell.*, vol.11, no.7, pp.674-693, July 1989.

[6] M. Vetterli and J. Kovačević, *Wavelets and Subband Coding*, Prentice Hall, Englewood Cliffs, NJ, 1995.

[7] P.H. Westerink, J. Biemond, and D.E. Boekee, "An optimal bit allocation algorithm for sub-band coding", *Proc. Int. Conf. on Acous., Speech and Signal Process.*, Dallas, Texas, pp. 1378-1381, April 1987.

[8] Y. Yoo, A. Ortega, and B. Yu, "Image Subband Coding using Progressive Classification and Adaptive Quantization" preprint, 1997.

Figure 4. Comparing results of various denoising methods, for *barbara* corrupted by noise $\sigma = 25$. From left to right, top to bottom: original, noisy observation, adaptive thresholding in DWT basis (*AdaptDWT*), uniform thresholding in DWT basis (*OrcUnifDWT*), spatial thresholding in overcomplete expansion (*AdaptNS*), and uniform thresholding in overcomplete expansion (*OrcUnifNS*). This figure can also be found at `http://www-wavelet.eecs.berkeley.edu/~grchang/icip98SpatialDenoise.pgm` .