Denoising Through Wavelet Shrinkage: An Empirical Study

Imola K. Fodor and Chandrika Kamath
Center for Applied Scientific Computing
Lawrence Livermore National Laboratory
P.O. Box 808, L-560, Livermore, CA 94551
fodor1@llnl.gov and kamath2@llnl.gov

July 27, 2001

Abstract

Techniques based on thresholding of wavelet coefficients are gaining popularity for denoising data. The idea is to transform the data into the wavelet basis, where the "large" coefficients are mainly the signal, and the "smaller" ones represent the noise. By suitably modifying these coefficients, the noise can be removed from the data. In this paper, we evaluate several two-dimensional denoising procedures using test images corrupted with additive Gaussian noise. We consider global, level-dependent, and subband-dependent implementations of these techniques. Our results, using the mean squared error as a measure of the quality of denoising, show that the SureShrink and the BayesShrink methods consistently outperform the other wavelet-based techniques. In contrast, we found that a combination of simple spatial filters led to images that were grainier with smoother edges, though the error was smaller than in the wavelet-based methods. [1]

Keywords: Image processing, denoising, wavelet thresholding, wavelet shrinkage, spatial filters.

[1] Submitted to the Journal of Electronic Imaging

1 Introduction

With sensors becoming ubiquitous and computers becoming more powerful, scientists are collecting and analyzing data at an ever increasing pace. In many fields such as astronomy, medical imaging, and computer vision, the data that is collected is often noisy, either as a result of the data acquisition process or due to natural phenomena such as atmospheric disturbances. This noise must be removed from the data before it can be analyzed.

Removing the noise from data can be considered as the process of constructing optimal estimates of the unknown signal from the available noisy data. There are several different ways in which this denoising can be done. In this paper, we investigate wavelet-based techniques for denoising, focusing on shrinkage methods. The basic idea behind these techniques is to use wavelets to transform the data into a different basis, where "large" coefficients correspond to the signal, while "small" ones represent mostly noise. The wavelet coefficients are suitably modified and the denoised data is obtained by an inverse wavelet transform of the modified coefficients.

In our work, we consider two dimensional versions of methods that were originally developed for one-dimensional signals in [1, 2, 3, 4, 5] and compare them to the method proposed for images in [6]. Using decimated wavelet transforms, and the mean squared error optimality criterion, we evaluate the different methods on test images corrupted with additive Gaussian white noise. Our goal is to address several issues. First, we want to better understand the sensitivity of the different methods to the choice of wavelet filters, the number of multiresolution levels, and the values of the parameters in each method. Second, we want to identify techniques that perform well across a variety of images and noise levels. Third, in contrast with other work, we want to explore the effect of different ways of modifying the wavelet coefficients for each method. By calculating and applying the modifications either globally, in a level-dependent manner, or in a subband-dependent manner, we hope to fine-tune the wavelet denoising to an image. Finally, we want to compare and contrast these wavelet-based techniques with the more traditional approaches based on spatial filters. Our goal is to complement the extensive theoretical and algorithmic work presented in the literature with a more practical, implementation-oriented comparison that would guide a practitioner in the choice of a method.

This paper is organized as follows. Section 2 gives a brief introduction to various denoising methods, followed by a detailed description of denoising through the shrinkage of wavelet coefficients. We describe the options available for such techniques and the different methods used to implement each option. Section 3 reports the results of the wavelet-based denoisers on test images with varying levels of additive white Gaussian noise. In Section 4, we compare the wavelet-based techniques to the more traditional approaches to denoising based on spatial filters. Finally, in Section 5, we conclude with a summary and our plans for future work.

2 Techniques for Removing Noise from Data

Spatial filters have long been used as the traditional means of removing noise from images and signals [7]. These filters usually smooth the data to reduce the noise, but, in the process, also blur the data. In the last decade, several new techniques have been developed that improve on spatial filters by removing the noise more effectively while preserving the edges in the data. Some of these techniques borrow ideas from partial differential equations and computational fluid dynamics such as level set methods [8, 9], total variation methods [10, 11], non-linear isotropic and anisotropic diffusion [12, 13], and essentially non-oscillatory (ENO) schemes [14]. Other techniques combine impulse removal filters with local adaptive filtering in the transform domain to remove not only white and mixed noise, but also their mixtures [15]. A different class of methods exploits the decomposition of the data into the wavelet basis and shrinks the wavelet coefficients to denoise the data [1, 2, 3, 4, 5, 6, 16, 17]. While this is typically done using the more memory efficient decimated wavelet transforms, it is well known that the use of non-decimated transforms minimizes the artifacts in the denoised data [18, 19]. Other authors have combined wavelets with Hidden Markov models and spatially adaptive methods [20, 21, 22, 23, 24, 25, 26], or used other basis functions such as ridgelets and curvelets [27, 28] that can be more effective than wavelets for images and higher dimensional data.

All these, and other, techniques have made image denoising a very active research area. However, what is lacking is a thorough comparison of the advantages and disadvantages of the different methods. The absence

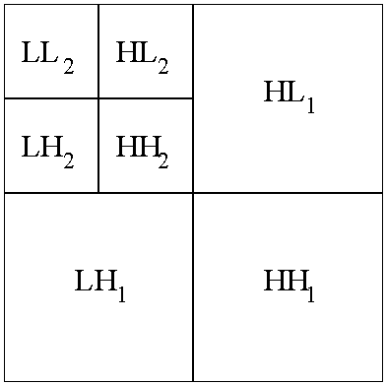

Figure 1: Wavelet decomposition subbands using a decimated transform with two multiresolution levels.

of such investigations, even within a single class of techniques, makes it very difficult for a practitioner to select an appropriate denoising scheme from the wealth of techniques that have been proposed in the literature. Our goal in this paper is to address this drawback in a small way by comparing and contrasting several of the different denoising methods that are based on the shrinkage of wavelet coefficients.

2.1 Denoising using Wavelet Shrinkage

The problem of denoising data can be stated as follows: given the zero-mean observation data $Y_{i,j}$ as a noisy realization of the signal $X_{i,j}$,

$$Y_{i,j} = X_{i,j} + \epsilon_{i,j}, \quad i = 1, \cdots, I, \ j = 1, \cdots, J, \ \epsilon_{i,j} \sim \mathcal{N}(0, \sigma^2), \tag{1}$$

construct an "optimal" estimate $\hat{X}_{i,j}$ of $X_{i,j}$ based on $Y_{i,j}$. In this work, we assume that the $\{\epsilon_{i,j}\}$s are independent from the signal and are independent and identically distributed Gaussian (normal) random variables with mean zero and variance σ^2. We also use the minimal mean squared error (MSE) to evaluate the optimality of the estimates. Let $\mathbf{Y}$, $\mathbf{X}$, and $\boldsymbol{\epsilon}$ denote the observed data, the noiseless data, and the error matrices in Eq. (1), respectively. Then, the three main steps of denoising using the wavelet coefficient shrinkage technique are as follows:

1. Calculate the wavelet coefficient matrix $\mathbf{w}$ by applying a wavelet transform $\mathbf{W}$ to the data:

$$\mathbf{w} = \mathbf{W}\ \mathbf{Y} = \mathbf{W}\ \mathbf{X} + \mathbf{W}\boldsymbol{\epsilon}, \tag{2}$$

2. Modify (i.e. threshold or shrink) the detail coefficients of $\mathbf{w}$ to obtain the estimate $\hat{\mathbf{w}}$ of the wavelet coefficients of $\mathbf{X}$:

$$\mathbf{w} \rightarrow \hat{\mathbf{w}}, \tag{3}$$

3. Inverse transform the modified coefficients to obtain the denoised estimate:

$$\hat{\mathbf{X}} = \mathbf{W}^{-1}\ \hat{\mathbf{w}}. \tag{4}$$

The number N of the wavelet coefficients $\mathbf{w}$ in Eq. (2) varies depending on the type of transform used. We focus on decimated transforms [29], where $N = IJ$, regardless of the number of multiresolution levels K, as it requires less memory than the undecimated transform. Fig. 1 displays the subbands of a two-level ($K = 2$) decimated decomposition. The coefficients on the first level are grouped into the vertical detail (LH_1), horizontal detail (HL_1), diagonal detail (HH_1), and smooth (LL_1) subbands. The smooth part is then similarly decomposed into the four second level subbands. The directions reflect the order in which the high-pass (H) and low-pass (L) filters of the wavelet transform are applied along the two dimensions of the original image.

The first step in denoising is to select a wavelet for the forward and inverse transformations $\mathbf{W}$ and $\mathbf{W^{-1}}$ in Eq. (2) and Eq. (4), respectively. We investigate well-known orthogonal and biorthogonal wavelets including the Daubechies family (daublets), the least asymmetric wavelet family (symmlets), the coiflet family (coiflets), and the B-spline and V-spline families [30]. These wavelets differ in their support, symmetry, and number of vanishing moments. In addition to a wavelet, we also need to select the number of multiresolution levels and an option for handling values near the edges of the image. We consider several boundary treatment rules [31], including periodic, symmetric, reflective, constant, and zero-padding.

The remainder of this section explains the details in the shrinkage (sometimes called thresholding) step in Eq. (3). Let w denote a single detail coefficient and $\hat{w}$ its shrunk (thresholded) version. Let λ be the threshold, $\delta_\lambda()$ denote the shrinkage function which determines how the threshold is applied to the data, and $\hat{\sigma}$ be an estimate of the standard deviation σ of the noise in Eq. (1). Then,

$$\hat{w} = \hat{\sigma}\ \delta_\lambda(w/\hat{\sigma}), \tag{5}$$

or

$$\hat{w} = \delta_\lambda(w), \tag{6}$$

depending on whether the threshold λ was determined assuming a unit noise scale $\sigma = 1$, in which case Eq. (5) applies, or an estimation of the actual noise was built-in into the method, in which Eq. (6) would be appropriate. Note that the noise estimate, the threshold, and the shrinkage function could depend on either the multiresolution level or the subband, though we have suppressed this dependence in our notation.

The denoising methods we consider differ in the choices for $\delta()$, λ and $\hat{\sigma}$, in Eq. (5) and (6). That is, we can obtain different denoisers by considering different

- shrinkage functions that determine how the threshold is applied (Section 2.2)
- noise estimates (Section 2.3), and
- shrinkage rules to determine the threshold λ (Section 2.4).

Since some shrinkage rules depend on the shrinkage functions and the noise estimates, we need to first select $\delta()$ and $\hat{\sigma}$ before we determine the λ.

For one-dimensional data, we can calculate the thresholds either globally, with one threshold for all the coefficients, or on a level-dependent basis, with K different thresholds for the K different dyadic levels. In two dimensions, in addition to these two possibilities, we can also calculate thresholds in a subband-dependent manner, and obtain $3K$ thresholds for the $3K$ detail coefficient subbands. While typical publications on denoising images [10, 6] consider either the level-dependent or the subband-dependent alternative in addition to the global implementation, we consider all the three different options. Next, we describe the different ways in which we can select the shrinkage functions, estimate the noise, and select the rules for denoising. More details can be found in [32], which is available on-line.

2.2 Shrinkage Functions

The shrinkage (thresholding) function determines how the thresholds are applied to the data. Fig. 2 displays the four thresholding functions we studied, scaled to the interval $[-1, 1]$. The x axis represents detail wavelet coefficients w in Eq. (2), and the y axis shows the corresponding thresholding function $\delta_\lambda(w)$. The dotted vertical lines indicate the values of the single threshold $\pm\lambda$ for the hard ($\delta^H_\lambda(w)$), soft ($\delta^S_\lambda(w)$), and garrote ($\delta^G_\lambda(w)$) functions. The semisoft function $\delta^{SS}_{\lambda_1,\lambda_2}(w)$ requires two thresholds, $\pm\lambda_1$ and $\pm\lambda_2$, represented by the four dotted vertical lines in its graph. If $I_{\{a\}}$ denotes the $\{0, 1\}$ indicator function, corresponding to $\{a = False, a = True\}$, the mathematical expressions for each of the shrinkage functions are

$$\delta^H_\lambda(w) = wI_{\{|w|>\lambda\}}, \tag{7}$$

$$\delta^S_\lambda(w) = \mathrm{sgn}(w)(|w| - \lambda)I_{\{|w|>\lambda\}}, \tag{8}$$

$$\delta^G_\lambda(w) = \left(w - \frac{\lambda^2}{w}\right)I_{\{|w|>\lambda\}}, \tag{9}$$

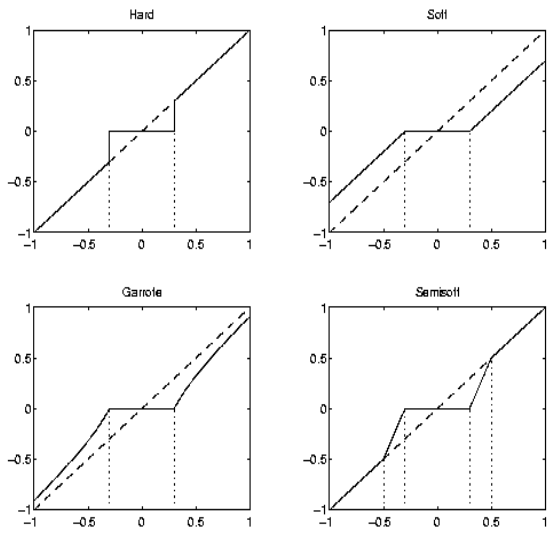

Figure 2: Shrinkage functions.

and

$$\delta^{SS}_{\lambda_1,\lambda_2}(w) = \left\{ \begin{array}{ll} 0, & |w| \leq \lambda_1 \\ \mathrm{sgn}(w)\frac{\lambda_2(|w|-\lambda_1)}{\lambda_2-\lambda_1}, & \lambda_1 < |w| \leq \lambda_2 \\ w, & |w| > \lambda_2. \end{array} \right\}. \tag{10}$$

2.3 Noise Estimates

Certain thresholds, as described in Section 2.4, are determined assuming a unit noise scale $\sigma = 1$. Therefore, in practice, the data must be standardized by an estimate of the noise scale, $\hat{\sigma}$, according to Eq. (5) before applying these thresholds. In some cases, there is prior knowledge about the noise distribution which can be used to obtain σ. However, in many situations, the noise must be estimated from the observed data. The options for estimating the noise scale include the choice of the functional form of the estimator and the choice of the detail coefficients to include in the estimation. We consider different estimation functions, including the sample standard deviation and the more robust median absolute deviation (MAD) suggested in [1]. We also consider noise estimates that are either global, level-dependent, or subband-dependent.

2.4 Shrinkage Rules

The shrinkage (thresholding) rules determine how the thresholds are calculated. Let λ denote a threshold. For convenience, the possible dependence of λ on the multiresolution level, $\{k, 1 \leq k \leq K\}$, or on the subband, $\{s, 1 \leq s \leq 3K\}$, is suppressed in the notation. Certain rules calculate the threshold independent of the shrinkage function, while others obtain different thresholds for different shrinkage functions. In addition, certain rules assume a unit noise scale, $\sigma = 1$; others do not. We indicate the assumptions of each method as we describe them in more detail.

- **Universal:**

  The *universal* rule was proposed in [2] as a global rule for one-dimensional signals. Regardless of the shrinkage function, for a signal of size N from a standard normal distribution $\mathcal{N}(0,1)$, the threshold is $\lambda = \sqrt{2\log N}$. With a noise estimate $\hat{\sigma}$, it is applied according to Eq. (5).

- **Minimizing the false discovery rate:**

Introduced in [5] for one-dimensional data, the *minFDR* rule determines the same global threshold for all shrinkage functions by keeping the expected value of the fraction of coefficients erroneously included in the reconstruction below a given fraction q. Given the N wavelet coefficients w_n, first it computes the p-values

$$p_n = 2[1 - \Phi(|w_n|/\hat{\sigma})], \tag{11}$$

where $\Phi()$ is the cumulative distribution function of the standard normal distribution, and $\hat{\sigma}$ is an estimate of the noise standard deviation. Then, it orders the p_n values as

$$p_{(1)} \leq p_{(2)} \leq \cdots \leq p_{(N)}. \tag{12}$$

Starting with $n = 1$, let m be the largest index n such that

$$p_{(n)} \leq \frac{n}{N}\, q. \tag{13}$$

The threshold is then obtained as

$$\lambda = \hat{\sigma}\Phi^{-1}\left(1 - \frac{p_{(m)}}{2}\right). \tag{14}$$

As $\hat{\sigma}$ is already factored in, the threshold is applied according to Eq. (6).

- **Top:**

  The *top* rule for one-dimensional signals [3] is a global method, independent of the shrinkage function. Given p as the fraction of the largest coefficients to keep, the threshold λ is set to be the $(1-p)$th quantile of the empirical distribution of the absolute values of the wavelet coefficients. It is applied to the coefficients using Eq. (6).

- **SURE:**

  For one-dimensional data, thresholds derived by minimizing Stein's Unbiased Risk Estimate (*SURE*) depend on the shrinkage function and on the multiresolution level [1]. The generalization to images can be achieved in either level- or subband-dependent manner. In the latter case, the threshold on subband s is

$$\lambda_s = \arg\min_{\lambda \geq 0} \mathrm{SURE}(\lambda, \mathbf{w}_s), \tag{15}$$

  where $\mathbf{w}_s$ denotes the detail coefficients from subband s, and $\mathrm{SURE}(\lambda, \mathbf{w}_s)$ denotes the corresponding Stein's unbiased estimate of the risk corresponding to a specific shrinkage function. For example, the threshold on subband s to be used with the soft shrinkage function, λ_s^S, is chosen as the value that minimizes $SURE^S(\lambda, \mathbf{w}_s)$,

$$\lambda_s^S = \arg\min_{\lambda \geq 0} SURE^S(\lambda, \mathbf{w}_s), \tag{16}$$

  where

$$\begin{aligned} SURE^S(\lambda, \mathbf{w}_s) &= N_s + \sum_{n=1}^{N_s} [\min(|w_n|, \lambda)]^2 \\ &\quad -2[(\#of\ w_n) : |w_n| \leq \lambda], \end{aligned} \tag{17}$$

  and N_s is the number of coefficients w_n in $\mathbf{w}_s$ [1, 32]. The threshold above assumes $\sigma = 1$. For data with non-unit variance, the coefficients are standardized by an appropriate $\hat{\sigma}$ estimate before calculating the threshold with Eq. (16). The level-dependent implementation is similar, except that instead of using the coefficients on a subband, one uses the coefficients on a level.

  It was shown in [1] that, in the case where the wavelet coefficient decomposition is sparse, a hybrid method combining the *universal* and the *SURE* thresholds is preferable over *SURE*. This hybrid method, when combined with the soft shrinkage function is referred to as *SureShrink* in the literature [1]. If

$$\frac{1}{N_s}\sum_{n=1}^{N_s}\left(\left(\frac{w_n}{\hat{\sigma}}\right)^2 - 1\right) \leq \frac{(\log_2 N_s)^{3/2}}{\sqrt{N_s}}, \tag{18}$$

then *SureShrink* uses the *universal* threshold, otherwise the *SURE* threshold is used for the coefficients on subband s.

Thresholds for the other shrinkage functions can also be derived. We worked out the details for hard thresholding, but in reporting the results, we simply call it *SURE* thresholding with hard shrinkage, instead of *WaveChop*, as suggested in [1].

- **Hypothesis Testing:**

  Introduced in [4] for one-dimensional signals, the *hypTest* rule calculates level-dependent thresholds independent of the thresholding function, based on testing the hypothesis that the wavelet coefficients at a given level are zero. In our extension to images, we distinguish between the level-dependent and the subband-dependent implementations.

  Following the notation in the description of the SURE rule for the subband-dependent version, and assuming that the N_s wavelet coefficients on subband s are normally distributed, we first find the largest of the squared wavelet coefficients on the subband, denoted by $w^2_{(N_s)}$, then compare it to the critical value

$$c^{\alpha}_{N_s} = \left\{ \Phi^{-1} \left[\frac{1}{2}((1-\alpha)^{1/N_s} + 1) \right] \right\}^2, \tag{19}$$

  where α is the pre-determined Type I error probability in the testing and $\Phi()$ is the cumulative distribution function of the standard normal density. If

$$\frac{w^2_{(N_s)}}{\hat{\sigma}^2} > c^{\alpha}_{N_s}, \tag{20}$$

  where $\hat{\sigma}$ is an estimate of the standard deviation of the noise, the null hypothesis of zero mean associated with the largest (in absolute value) coefficient is rejected, and so $w_{(N_s)}$ is retained as signal. Next, the process is iterated with the square of the second largest (in absolute value) wavelet coefficient $w^2_{(N_s-1)}$. If $w^2_{(N_s-1)}/\hat{\sigma}^2 > c^{\alpha}_{N_s-1}$, the procedure continues until at some point the mth largest (in absolute value) coefficient satisfies

$$\frac{w^2_{(m)}}{\hat{\sigma}^2} \leq c^{\alpha}_{m}. \tag{21}$$

  The threshold at subband s is then set as

$$\lambda_s = |w_{(m)}|, \tag{22}$$

  and is applied according to Eq. (6).

- **BayesShrink:**

  The *BayesShrink* rule of [6] uses a Bayesian mathematical framework for images to derive subband-dependent thresholds that are nearly optimal for soft thresholding. The formula for the threshold on a given subband s is:

$$\lambda_s = \frac{\hat{\sigma}^2}{\hat{\sigma}_X}, \tag{23}$$

  where $\hat{\sigma}^2$ is the estimated noise variance, and $\hat{\sigma}^2_X$ is the estimated signal variance on the subband considered. The noise variance is estimated as the median absolute deviation of the diagonal detail coefficients on level 1 (i.e. subband HH_1). The estimate of the signal standard deviation is

$$\hat{\sigma}_X = \sqrt{\max(\hat{\sigma}^2_Y - \hat{\sigma}^2, 0)}, \tag{24}$$

  where

$$\hat{\sigma}^2_Y = \frac{1}{N_s} \sum_{n=1}^{N_s} w_n^2 \tag{25}$$

is an estimate of the variance of the observations, with N_s being the number of the wavelet coefficients w_n on the subband under consideration. In case $\hat{\sigma}^2 \geq \hat{\sigma}_Y^2$, the threshold is set to $\lambda_s = \max(|w_n|)$, and all coefficients from the subband are set to zero. These thresholds are applied according to Eq. (6).

This method has been proposed for use with soft thresholding. We use the thresholds calculated via this procedure with other thresholding functions as well, but, in compliance with [6], we reserve the term *BayesShrink* for denoising with the soft shrinkage function.

It is clear from the description of the various shrinkage rules, that some of them require input parameters. We next describe our experimental results, including the empirical determination of the values of these parameters.

3 Experimental Results

We compared the various denoising methods described in Section 2 on several test images widely used in the image processing community. Here, we report the results only for the Lena image. Complete results for other test images can be found in [32]. Our experimental approach was as follows. First, we obtained the 512×512 pixel, noiseless, grayscale originals from http://www.image.cityu.edu.hk/imagedb/. We corrupted these images by adding Gaussian noise to the images according to Eq. (1), using $\sigma = 10, 20$, and 30. Fig. 3 (a) shows the original Lena image and Fig. 3 (b) shows the image with additive Gaussian noise of $\sigma = 20$. Next, we determined the parameters for the shrinkage rules as described in Section 3.1. Using these parameters, we applied the wavelet denoising methods, each in a global, level-dependent, and subband-dependent manner. We evaluated the quality of the denoising using the mean squared error (MSE) and the signal-to-noise-ratio (SNR) defined below. For a given estimate $\hat{X}(i,j)$ of $X(i,j)$, the MSE is

$$MSE = \frac{1}{IJ} \sum_{i=1}^{I} \sum_{j=1}^{J} (X(i,j) - \hat{X}(i,j))^2, \tag{26}$$

the corresponding normalized MSE is

$$MSE_n = \frac{\sum_{i=1}^{I} \sum_{j=1}^{J} (X(i,j) - \hat{X}(i,j))^2}{\sum_{i=1}^{I} \sum_{j=1}^{J} X(i,j)^2}, \tag{27}$$

and the SNR on dB scale [19] is

$$SNR = 10 \ \log_{10} \frac{1}{MSE_n}. \tag{28}$$

3.1 Selection of Parameters

As described in [32], we determined the "optimal" parameters for the *minFDR*, *top*, and *hypTest* denoising methods empirically, using the 512×512 pixel grayscale Lena test image with additive Gaussian noise with $\sigma = 10$, the symmlet8 wavelet, four multiresolution levels and periodic boundary treatment. We chose the symmlet8 wavelet as it is relatively symmetric, and has a reasonably compact support.

For the *minFDR* rule, we found the optimal parameters to be $q = 0.2$ for the global implementation, and $q = 0.3$ for the level and the subband-dependent implementations. The global implementation was superior to the level-dependent one, which, in turn, was superior to the subband-dependent version.

For the *top* procedure, we determined the best parameter as $p = 0.3$ for the global implementation, and as $p_1 = .15$, $p_2 = .4$, $p_3 = .8$, and $p_4 = .95$ on multiresolution levels one through four, respectively, for the level-dependent and the subband-dependent cases. The level-dependent method was the best, followed by the subband-dependent, followed by the global implementation. For the *top* rule with the semisoft shrinkage function, we calculated the thresholds by using two different p parameters. The optimal parameter pair for the global implementation was $\{p_1 = 0.1,\ p_2 = 0.01\}$. We obtained the best subband and level-dependent results with $\{p_1^{(1)} = .15, p_2^{(1)} = .1\}, \{p_1^{(2)} = .3, p_2^{(2)} = .1\}, \{p_1^{(3)} = .6, p_2^{(3)} = .2\}, \{p_1^{(4)} = .7, p_2^{(4)} = .3\}$,

where the superscripts indicate the multiresolution levels. The global implementation was the best, the subband-dependent was second, and the level-dependent was third.

For the *hypTest* rule, in accordance with [33], we found that an unusually high value, $\alpha = 0.9$, was the optimal. The subband-dependent implementation outperformed the level-dependent version, which outperformed the global method.

For the methods that required an estimate of σ, we used the MAD of the detail coefficients on the HH_1 subband [1], as it was the most robust of the alternatives we tried. A typical such estimate in the case when $\sigma = 10$ is $\hat{\sigma} = 10.52$.

3.2 Comparison of Wavelet-based Techniques

In Table 1, we present the results of denoising for the Lena image. The three main columns report the MSE and SNR values corresponding to the three different noise levels. The first row reports the MSE and SNR values for the noisy images, enabling evaluation of the denoised results. The thresholding rules are prefixed with either $S\_$ or $P\_$ to indicate whether the thresholds were calculated globally (i.e. *S*ingle threshold), or dependent on the level or subband (i.e. *P*yramid of thresholds). For all methods, we considered both level- and subband-dependent implementations, but report only the better of the two approaches for each method. The boldfaced entry in each column indicates the best method for the corresponding noise level. Fig. 3 displays a few examples of the denoised results.

It is clear from Table 1, that there is a wide range of variation in the quality of the denoised images. For example, for the $\sigma = 20$ image, although all methods decrease the $MSE = 399.50$ of the noisy image, the denoised MSE values range from the worst of 351.63 to the best of 61.59. Choosing the right method therefore has a large effect on the results.

We next summarize our observations based on the experimental results reported here, and in the more detailed study [32]. Except for the different quantities involved, the main conclusions reached for the Lena image are valid for the other images considered in the detailed study. In the following, recall that *SureShrink* refers to the subband-dependent P_SURE thresholding with soft shrinkage and *BayesShrink* refers to the subband-dependent P_Bayes method with soft shrinkage.

- **Influence of Shrinkage Function:**

  As the values in Table 1 indicate, in most cases, soft shrinkage was superior to garrote shrinkage, which, in turn, was superior to hard shrinkage. An important exception occurred with the *universal* rule, where both the hard and garrote functions gave better estimates than the soft function, regardless of the noise level. In the case of the *top* rule, subband-dependent semisoft shrinkage was always inferior to the corresponding soft shrinkage results, but the global semisoft implementation led to values comparable to those obtained with the global soft thresholding for $\sigma = 10$ and $\sigma = 20$.

  We conclude that the choice of the shrinkage function strongly influences the results, and that the soft shrinkage function is preferred to either the garrote, hard, or semisoft functions. We note that, for statistical reasons, the authors in [6] only consider soft shrinkage.

- **Influence of Shrinkage Rule:**

  The range of values in Table 1 indicates that the shrinkage rule strongly affects the outcome of the denoising operation. In most cases, the pyramidal implementations of the rules resulted in better estimates than the global implementations, regardless of the noise level. For the *minFDR* method, the two implementations led to very similar results. For the *hypTest* method, the results depended on the noise level: the pyramidal version was superior for $\sigma = 10$, but it was inferior for $\sigma = 20$ and $\sigma = 30$.

  Despite the global implementation proposed in [2], we found that subband-dependent *universal* thresholding outperformed its global version, regardless of the shrinkage function and the noise level.

  In all cases, we found that *SureShrink* was the best method. *BayesShrink* was the second best in all cases but one ($\sigma = 10$), where the level-dependent *top* method was slightly better. Since we did not observe this consistently across different images and noise levels, we exclude the *top* rule from the list of best denoisers. However, we believe that more data-adaptive tuning is needed to choose the optimal parameter for the *top* method than what we proposed in Section 3.1. We also observe that in most

of the examples considered in [6], the *BayesShrink* procedure proposed by the authors outperforms *SureShrink*.

- **Influence of Noise:**

  Our main conclusion that *SureShrink* and *BayesShrink* were the best denoisers overall is not sensitive to the amount of noise in the images. We do, however, stress that it is important to use a robust estimator of the noise, such as the MAD described in Section 2.3.

- **Influence of Wavelets:**

  In our experiments with denoising, we found that the choice of wavelet, the number of multi-resolution levels, and the boundary treatment rule had little effect on the results. This observation agrees with [6]. The results in Table 1 were obtained using the symmlet12 wavelet with 12 coefficients [30] with $K = 3$ multiresolution levels and periodic boundary treatment. We performed the same analyses using different wavelets, different numbers of levels, and different boundary extensions [32]. The ordering of the methods, as measured by their MSE values, remained the same for the alternatives we tried.

  In our study, the biorthogonal wavelets fared worse, as measured by the MSE, than the orthogonal symmlets. Because of their symmetry, the biorthogonal wavelets are claimed to lead to fewer visual artifacts in reconstructed images [10]. In our experiments, however, images denoised with the nearly-symmetric orthogonal symmlets are visually comparable to those obtained with biorthogonal wavelets.

- **VisuShrink Compared to SureShrink:**

  The authors in [2] define the term *VisuShrink* to refer to global soft shrinkage with the *universal* threshold for one-dimensional signals, because it leads to visually pleasing results. However, for two-dimensional images, just as the authors in [10], we found that *SureShrink* yielded much better results than the *VisuShrink* procedure, both in terms of MSE and visual quality. In fact, we found that even the global hard thresholding with the *universal* threshold outperformed *VisuShrink*. Though far from optimal, this global hard thresholding with the *universal* threshold is sometimes used as a benchmark in measuring denoising performance. Fig. 3 (d) displays the result of this method on Lena. In comparison, *SureShrink* in panel (c) results in much superior denoising.

In conclusion, we found that in our experiments, *SureShrink* and *BayesShrink* were the best denoisers among the ones we studied. They yielded similar results and consistently outperformed the other methods in all but one case (the Lena image with $\sigma = 10$, where the level-dependent *top* method with soft shrinkage resulted in a slightly smaller MSE than *BayesShrink*). In most cases, *SureShrink* had slightly smaller MSE values than *BayesShrink*, but the differences were small enough to be just random fluctuations, and are in agreement with [6]. However, some users might prefer the *BayesShrink* method because of its simplicity.

We realize that, for methods with input parameters, we could have fine-tuned the parameters to the image, and reduced the MSE further, making the methods more competitive. However, this need for fine-tuning could be considered a drawback of these methods, especially since the two top contenders, *SureShrink* and *BayesShrink* achieved nearly optimal performance without any tuning of parameters.

4 Comparison with Spatial Filters

While wavelet-based denoising techniques are certainly a powerful tool for image restoration, our study would be incomplete without a comparison with the more traditional approaches based on the use of spatial filters [7, 34]. In this section, we compare the effectiveness of denoising using several linear and non-linear filters applied either by themselves or in combination with other spatial filters. Our choice of filters is listed below, where the filter size is indicated in parenthesis:

- Mean filters $(3 \times 3, 5 \times 5)$
- Gaussian filters $(3 \times 3, 5 \times 5)$

- Scaled unsharp masking filters ($3 \times 3, 5 \times 5$). Given the real number β, these filters calculate $(1.0 + \beta)$original_image $- (\beta)$mean_filtered_image. In our experiments, $\beta = -0.8$ gave relatively good results.
- Alpha-trimmed mean filters ($3 \times 3, 5 \times 5$) with a trim size of either 1 or 2. The trim size is the number of smallest and largest pixels that are excluded in the calculation of the mean.
- Median filters ($3 \times 3, 5 \times 5$)
- Mid-point filters ($3 \times 3, 5 \times 5$). The value calculated is the average of the minimum and maximum within the filter mask.
- Minimum mean squared error filters ($3 \times 3, 5 \times 5$)

For filters requiring an estimate of the noise variance, we used the following algorithm. First, we subtracted a (3×3) mean-filtered image from the original. Next, we calculated the standard deviation, *sd*, of the resulting image, and dropped all the pixels whose absolute values were larger than *sd*. Finally, we used the standard deviation of the remaining pixels as the estimate of the standard deviation σ of the noise in the image. For the Lena image with $\sigma = 10$, a typical estimate obtained with this method is $\hat{\sigma} = 10.95$.

The spatial filter-based findings for the Lena image are given in Table 2, the bold entries indicating the best results for the different noise levels. We used periodic boundary treatment to handle values near the edges of the image. Examples of the denoised images are given in Fig. 3.

The authors in [6] show empirically that the results of the best possible linear filtering, using the Wiener filter, are inferior to the results obtained with *SureShrink*. Comparing the MSE values in Table 2 to the corresponding values in Table 1, however, indicates that combinations of spatial filters can be very competitive with wavelet-based denoising techniques. For $\sigma = 20$ and $\sigma = 30$, the best spatial-filtered images, using the 5×5 minimum MSE filter followed by a 3×3 mean filter, have smaller errors than the best wavelet-denoised images.

Spatial filters are very simple to implement and are computationally faster than wavelet-based methods as they require far less computation in many cases. However, a comparison of the images indicates that spatial filters often result in grainier images than the ones obtained from wavelet techniques. Unless special care is taken near the edges, they also tend to smooth the edges in the image. On the other hand, wavelet-based approaches sometime create noticeable artifacts that can substantially degrade the image.

5 Summary

In this paper, we evaluated several denoising methods on test images corrupted with white Gaussian noise. We considered an extensive set of techniques based on statistical thresholding of wavelet coefficients as well as the more traditional approaches using spatial filters.

Based on our experiments, we conclude that *SureShrink* and *BayesShrink* are the best wavelet-based denoising methods for the types of images we considered, among the methods we considered. None of the other wavelet-based procedures that we examined achieved lower error rates, as measured by the MSE, than these two techniques. When we consider simplicity of implementation along with the denoising performance, we found *BayesShrink* to be the best procedure.

For completeness, we also compared the wavelet-based denoisers with various spatial filter-based methods. Our results indicate that on an case-by-case basis, it is often possible to find a denoiser based on combinations of spatial filters that is superior to the best wavelet-based denoiser. In most, but not all, of those cases, the optimal method is given by the 5×5 Min-MSE filter followed by the 3×3 Gaussian filter.

As we have mentioned, there are several alternative methods for denoising image and other data that have been proposed in the last few years. We plan to complement our study of wavelet-based shrinkage techniques by exploring some of these newer techniques as well.

6 Acknowledgments

UCRL-JC-144258. This work was performed under the auspices of the U.S. Department of Energy by University of California Lawrence Livermore National Laboratory under contract No. W-7405-Eng-48.

References

[1] D.L. Donoho and I. M. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage," *Journal of the American Statistical Association*, vol. 90, no. 432, pp. 1200–1224, December 1995.

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

[3] A. Bruce and H. Gao, *Applied Wavelet Analysis with S-PLUS*, Springer Verlag, 1996.

[4] R.T. Ogden, *Essential Wavelets for Statistical Applications and Data Analysis*, Birkhäuser, 1997.

[5] B. Vidakovic, *Statistical Modeling by Wavelets*, Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., 1999.

[6] S.G. Chang, B. Yu, and M. Vetterli, "Adaptive wavelet thresholding for image denoising and compression," *IEEE Trans. Image Processing*, vol. 9, pp. 1532–1546, 2000.

[7] A.R. Weeks, *Fundamentals of Electronic Image Processing*, SPIE/IEEE Series on Imaging Science and Engineering. SPIE Optical Engineering Press and IEEE Press, 1996.

[8] R. Malladi and J. Sethian, "A unified approach to noise removal, image enhancement, and shape recovery," *IEEE Trans. Image Processing*, vol. 5, pp. 1154–1168, 1996.

[9] J.A. Sethian, *Level Set Methods and Fast Marching Methods: Evolving Interfaces in Computational Geometry, Fluid Mechanics, Computer Vision and Materials Science*, Cambridge University Press, 1999.

[10] A. Chambolle, R.A. DeVore, N. Lee, and B.J. Lucier, "Nonlinear wavelet image processing: variational problems, compression, and noise removal through wavelets," *IEEE Trans. Image Processing*, vol. 7, pp. 319–335, 1998.

[11] T. Chan and H. Zhou, "Optimal construction of wavelet coefficients using total variation regularization in image compression," Tech. Rep. Computational and Applied Mathematics 00-27, UCLA, 2000.

[12] M. Black, G. Sapiro, D. Marimont, and D. Heeger, "Robust anisotropic diffusion," *IEEE Trans. Image Processing*, vol. 7, pp. 421–432, 1998.

[13] J. Weickert, B.M. ter Haar Romeny, and M. Viergever, "Efficient and reliable schemes for nonlinear diffusion filtering," *IEEE Trans. Image Processing*, vol. 7, pp. 398–410, 1998.

[14] T. Chan and H. Zhou, "Adaptive ENO-wavelet transforms for discontinuous functions," Tech. Rep. Computational and Applied Mathematics 99-21, UCLA, 1999.

[15] K. Egiazarian, J. Astola, M. Helsingius, and P. Kuosmanen, "Adaptive denoising and lossy compression of images in transform domain," *Journal of Electronic Imaging*, vol. 8, no. 3, pp. 233–245, July 1999.

[16] J.B. Weaver, X. Yansun, Jr. D.M. Healy, and L.D. Cromwell, "Filtering noise from images with wavelet transforms," *Magnetic Resonance in Medicine*, vol. 24, pp. 288–195, 1991.

[17] R.A. DeVore and B.J. Lucier, "Fast wavelet techniques for near-optimal processing," IEEE Military Communications Conference, 1992, pp. 48.3.1–48.3.7, New York.

[18] R.R. Coifman and D.L. Donoho, "Translation-invariant de-noising," in *Wavelets and Statistics*, A. Antoniadis and G. Oppenheim, Eds., Lecture Notes in Statistics, pp. 125–150. Springer-Verlag, 1995.

[19] J.-L. Starck, F. Murtagh, and A. Bijaoui, *Image Processing and Data Analysis. The Multiscale Approach*, Cambridge University Press, 1998.

[20] M.S. Crouse, R. D. Nowak, and R. G. Baraniuk, "Wavelet-based signal processing using hidden Markov models," *IEEE Transactions on Signal Processing*, vol. 46, no. 4, pp. 886–902, 1998.

[21] J.K. Romberg, H. Choi, and R. G. Baraniuk, "Shift-invariant denoising using wavelet-domain hidden Markov trees," in *Conference Record of The Thirty-Third Asilomar Conference on Signals, Systems and Computers*, October 1999.

[22] M.K. Mihcak, "Low-complexity image denoising based on statistical modeling of wavelet coefficients," *IEEE Signal Processing Letters*, vol. 6, no. 12, pp. 300–303, Dec 1999.

[23] S.G. Chang, B. Yu, and M. Vetterli, "Spatially adaptive wavelet thresholding based on context modeling for image denoising," *IEEE Trans. Image Processing*, vol. 9, pp. 1522–1531, 2000.

[24] Y. Wan and R.D. Nowak, "A wavelet-based statistical model for image restoration," in *Proc. IEEE International Conference on Image Processing - ICIP'01, Thessaloniki, Greece*, October 2001, to appear.

[25] G.L Fan and X.-G. Xia, "Image denoising using local contextual hidden Markov model in the wavelet domain," *IEEE Signal Processing Letters*, vol. 8, no. 5, pp. 125–128, May 2001.

[26] J. Liu and P. Moulin, "Complexity-regularized image denoising," *IEEE Trans. on Image Processing*, June 2001, to appear.

[27] E.J. Candès, *Ridgelets: Theory and Applications*, Ph.D. thesis, Stanford University, 1998.

[28] E.J. Candès and D.J. Donoho, "Curvelets, multiresolution representation, and scaling laws," in *Wavelet Applications in Signal and Image Processing VIII*, A. Aldroubi, A. F. Laine, and M. A. Unser, Eds. SPIE, 2000, vol. 4119.

[29] S.G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 11, pp. 674–693, 1989.

[30] I. Daubechies, *Ten Lectures on Wavelets*, SIAM, 1992.

[31] M. Unser, "A practical guide to the implementation of the wavelet transform," in *Wavelets in Medicine and Biology*, A. Aldroubi and M. Unser, Eds., pp. 37–73. CRC Press, 1996.

[32] I.K. Fodor and C. Kamath, "Wavelet denoising of images: A comparison of methods," Technical report UCRL-JC-142357, Lawrence Livermore National Laboratory, 2001, http://www.llnl.gov/casc/sapphire/.

[33] C. Cai and P. de B. Harrington, "Different discrete wavelet transforms applied to denoising analytical data," *J. Chem. Inf. Comput. Sci.*, vol. 38, pp. 1161–1170, 1998.

[34] S. Umbaugh, *Computer Vision and Image Processing: A Practical Approach using CVIPtools*, Prentice Hall, 1998.

Figure 3: Denoising results for the Lena image. (a) Original image. (b) Noisy image, $\sigma = 20$, $MSE = 399.50$. (c) *SureShrink*, $MSE = 61.59$. (d) Global *universal* rule with hard thresholding, $MSE = 103.95$. (e) Minimum mean squared error (5×5) followed by Gaussian (3×3) filter, $MSE = 56.80$. (f) Minimum mean squared error (5×5) followed by mean (3×3) filter, $MSE = 64.80$.

Table 1: Wavelet-based results for the Lena image, the symmlet12 wavelet, three multiresolution levels and periodic boundary treatment.

		$\sigma = 10$		$\sigma = 20$		$\sigma = 30$	
	Rule	**MSE**	**SNR**	**MSE**	**SNR**	**MSE**	**SNR**
Noisy image		99.53	13.62	399.50	7.58	894.66	4.08
Soft	S_Universal	86.57	14.22	136.52	12.25	169.40	11.31
	S_MinFDR	33.26	18.38	81.13	14.51	143.89	12.02
	S_Top	33.35	18.37	91.46	13.99	177.24	11.11
	S_HypTest	33.50	18.35	103.12	13.47	261.45	9.42
	S_SURE	33.70	18.32	74.91	14.85	113.01	13.07
	S_Bayes	39.09	17.68	75.77	14.80	115.18	12.98
	P_Universal	76.45	14.76	123.54	12.68	156.90	11.64
	P_MinFDR	33.59	18.34	81.80	14.47	143.74	12.02
	P_Top	29.92	18.84	75.06	14.84	141.46	12.09
	P_HypTest	31.83	18.57	112.41	13.09	292.43	8.94
	P_SURE	**29.23**	**18.94**	**61.59**	**15.70**	**91.34**	**13.99**
	P_Bayes	30.26	18.79	63.33	15.58	92.74	13.93
Hard	S_Universal	55.98	16.12	103.95	13.43	142.44	12.06
	S_MinFDR	75.40	14.83	282.55	9.09	622.14	5.66
	S_Top	80.20	14.56	317.27	8.58	709.56	5.09
	S_HypTest	72.39	15.00	339.18	9.29	810.56	4.51
	S_SURE	70.23	15.13	257.76	9.49	562.66	6.10
	S_Bayes	92.38	13.94	240.58	9.79	255.64	9.52
	P_Universal	49.55	16.65	93.58	13.89	128.61	12.51
	P_MinFDR	75.11	14.84	280.66	9.12	618.50	5.69
	P_Top	68.56	15.24	263.28	9.39	584.97	5.93
	P_HypTest	76.71	14.75	351.63	8.14	829.56	4.41
	P_SURE	70.24	15.13	257.81	9.49	561.53	6.10
	P_Bayes	48.26	16.76	101.29	13.54	152.54	11.76
Garrote	S_Universal	69.81	15.16	121.28	12.76	158.12	11.61
	S_MinFDR	37.70	17.83	115.52	12.97	234.12	9.90
	S_Top	41.09	17.46	143.96	12.02	311.00	8.67
	S_HypTest	36.12	18.02	169.95	11.30	463.92	6.93
	S_Bayes	58.01	15.96	93.61	13.89	112.14	13.10
	P_Universal	60.98	15.75	107.77	13.27	143.52	12.03
	P_MinFDR	37.60	17.85	114.75	13.00	231.91	9.95
	P_Top	34.79	18.18	111.59	13.12	234.01	9.91
	P_HypTest	38.36	17.76	190.23	10.81	511.85	6.51
	P_Bayes	34.37	18.24	71.35	15.06	103.66	13.44
SemiSoft	S_Top	33.09	18.40	91.46	13.99	200.92	10.57
	P_Top	56.95	16.04	214.62	10.28	475.91	6.82

Table 2: Filter-based results for the Lena image with periodic boundary treatment.

	$\sigma = 10$		$\sigma = 20$		$\sigma = 30$	
Image	**MSE**	**SNR**	**MSE**	**SNR**	**MSE**	**SNR**
Noisy image	99.53	13.62	399.50	7.58	894.66	4.08
Mean (3)	42.67	17.30	76.34	14.77	130.46	12.44
Mean (5)	82.07	14.46	93.91	13.87	113.59	13.05
Gaussian (3)	32.95	18.42	95.25	13.81	196.92	10.66
Gaussian (5)	45.58	17.01	67.09	15.33	101.58	13.53
Scaled Unsharp Masking (3)	34.84	18.18	79.13	14.62	150.75	11.82
Scaled Unsharp Masking (5)	57.79	15.98	81.32	14.50	119.63	12.82
Alpha Trimmed Mean (3,1)	39.12	17.68	76.11	14.78	134.52	12.31
Alpha Trimmed Mean (5,2)	74.20	14.89	88.48	14.13	110.41	13.17
Median (3)	40.16	17.56	94.50	13.84	178.17	11.09
Median (5)	59.31	15.87	84.41	14.33	120.39	12.79
Mid-point (3)	78.76	14.64	131.29	12.42	222.31	10.13
Mid-point (5)	178.00	11.09	196.15	10.67	249.49	9.63
Min-MSE (3)	70.67	15.11	192.53	10.75	389.93	7.69
Min-MSE (5)	39.35	17.65	91.97	13.96	163.36	11.47
S-Unsharp (3), Mean (3)	50.50	16.57	69.00	15.21	98.65	13.66
Mean (3), Mean (3)	54.07	16.27	70.67	15.11	97.35	13.72
Min-MSE (5), Mean (3)	44.95	17.07	64.80	15.48	89.44	14.08
Min-MSE (5), Gaussian (3)	**31.78**	**18.58**	**56.80**	**16.06**	**88.52**	**14.13**
S-Unsharp (3), Gaussian (3)	41.56	17.41	66.99	15.34	107.81	13.27
Gaussian (3), Gaussian (3)	37.09	17.91	68.64	15.23	119.50	12.83