Feature-Based Wavelet Shrinkage Algorithm for Image Denoising

Eric J. Balster, *Member, IEEE*, Yuan F. Zheng, *Fellow, IEEE*, and Robert L. Ewing, *Senior Member, IEEE*

***Abstract*—A selective wavelet shrinkage algorithm for digital image denoising is presented. The performance of this method is an improvement upon other methods proposed in the literature and is algorithmically simple for large computational savings. The improved performance and computational speed of the proposed wavelet shrinkage algorithm is presented and experimentally compared with established methods. The denoising method incorporated in the proposed algorithm involves a two-threshold validation process for real-time selection of wavelet coefficients. The two-threshold criteria selects wavelet coefficients based on their absolute value, spatial regularity, and regularity across multiresolution scales. The proposed algorithm takes image features into consideration in the selection process. Statistically, most images have regular features resulting in connected subband coefficients. Therefore, the resulting subbands of wavelet transformed images in large part do not contain isolated coefficients. In the proposed algorithm, coefficients are selected due to their magnitude, and only a subset of those selected coefficients which exhibit a spatially regular behavior remain for image reconstruction. Therefore, two thresholds are used in the coefficient selection process. The first threshold is used to distinguish coefficients of large magnitude and the second is used to distinguish coefficients of spatial regularity. The performance of the proposed wavelet denoising technique is an improvement upon several other established wavelet denoising techniques, as well as being computationally efficient to facilitate real-time image-processing applications.**

***Index Terms*—Image denoising, selective wavelet shrinkage, two-threshold criteria.**

I. Introduction

THE recent advancement in multimedia technology has promoted an enormous amount of research in the area of image and video processing. Included in the many image and video processing applications, such as compression, enhancement, and target recognition, are preprocessing functions for noise removal. Noise removal is one of the most common and important processing steps in many image and video systems.

Because of the importance and commonality of preprocessing in most image and video systems, there has been an enormous amount of research dedicated to the subject of noise removal, and many different mathematical tools have been proposed. Variable coefficient linear filters [6], [19], [22], [29], adaptive nonlinear filters [11], [18], [20], [31], DCT-based solutions [13], cluster filtering [28], genetic algorithms [27], fuzzy logic [14], [24], etc., have all been proposed in the literature.

The wavelet transform has also been used to suppress noise in digital images. It has been shown that the reduction of absolute value in wavelet coefficients is successful in signal restoration [16], [17], [21], [30]. This process is known as wavelet shrinkage. Preliminary methods of wavelet shrinkage predict the contribution of wavelet coefficients based on their magnitude [26], [30], and others predict the contribution based on intrascale dependencies of wavelet coefficients [4], [8], [15], [17]. More recent denoising methods are based on both intra- and interscale coefficient dependencies [7], [10], [12], [16], [21]. Also, in addition to the calculation of a coefficient's contribution, there are two basic approaches to modifying the coefficient. We define them as *probabilistic* wavelet shrinkage and *selective* wavelet shrinkage.

The difference between probabilistic wavelet shrinkage methods of [16] and [21] and the selective wavelet shrinkage methods of [17] and [30] is in the modification of the wavelet coefficients. In the first method, the level of reduction of coefficient magnitude is continuous between 0 and 1. In other words, the magnitude of the wavelet coefficient is reduced by the probability of its contribution to the overall quality of the image. The second method uses a binary method where the reduction of coefficient magnitude is either 0 or 1, i.e., coefficients are either selected or removed. both methods have to evaluate the contribution of the wavelet coefficients to determine the modification, but the first method usually involves more computation not necessarily resulting in better performance, as is shown in this paper.

Mallat and Hwang prove the successful removal of noise in signals via selective wavelet shrinkage method based on Lipschitz (Hölder) exponents [17]. The Hölder exponent is a measure of regularity in a signal, and it may be approximated by the evolution of wavelet coefficient ratios across scales. Thus, this regularity metric used in selecting those wavelet coefficients which are to be used in reconstruction, and those which are not. Although this fundamental work in image denoising is successful in the removal of noise, its application is broad and not focused on image noise removal, and the results are not optimal.

Malfait and Roose use the probabilistic shrinkage method by applying a Bayesian probabilistic formulation, and modeling the wavelet coefficients as Markov random sequences [16]. This method is focused on image denoising and its results are an improvement upon [17]. The Lipschitz (Hölder) exponents

Manuscript received August 18, 2003; revised December 1, 2004. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Eli Saber.

E. J. Balster and R. L. Ewing are with the Air Force Research Laboratory, Wright-Patterson Air Force Base, Dayton, OH 45433-7334 USA.

Y. F. Zheng is with The Ohio State University, Columbus, OH 43210 USA.

Digital Object Identifier 10.1109/TIP.2005.859385

are roughly approximated by the evolution of coefficient values across scales, i.e.,

$$m_{l,n} = \frac{1}{p-l}\sum_{k=l}^{p-1}\left|\frac{\lambda_{k+1,n}}{\lambda_{k,n}}\right|$$

where $m_{l,n}$ is the approximated Hölder exponent of position n of scale l and $\lambda_{k,n}$ is the wavelet coefficient of scale k and position n. The rough approximation is refined by assuming that the coefficient values are well modeled as a Markov chain, and the probability of a coefficient's contribution to the image can be well approximated by the Hölder exponents of neighboring coefficients. Coefficients are then assigned binary labels $v_{k,n}$ of position n depending on their predicted retention for reconstruction ($v_{k,n} = 1$), or predicted removal ($v_{k,n} = 0$). The binary labels are then randomly and iteratively switched until $P(V|M)$ is maximized, where $v_{k,n} \in V$ and $m_{k,n} \in M$. The coefficients are modified by $\lambda_{k,n}^{\text{new}} = \lambda_{k,n}P(v_{k,n} = 1|M)$, and the denoised image is formed by the inverse wavelet transform of the modified coefficients. Each coefficient is reduced in magnitude depending on the probable contribution to the image, i.e., $P(v_{k,n} = 1|M)$.

Later, Pizurica, *et al.* [21] continued on the work done by [16] by using a different approximation of the Hölder exponent given by

$$\rho_{l,n} = \frac{1}{p-l}\sum_{k=l}^{p-1}\left|\frac{I_{k+1,n}}{I_{k,n}}\right|$$

where

$$I_{k,n} = \sum_{t\in C(k,n)}|\lambda_{k,t}|.$$

$\rho_{l,n}$ is the approximation of the Hölder exponent, and $C(k,n)$ is the set of coefficients surrounding $\lambda_{k,n}$. This work applies the same probabilistic model as in [16], using the new approximation of the Hölder exponent. Coefficients are assigned binary labels, $v_{k,n}$, depending on their predicted retention for reconstruction ($v_{k,n} = 1$), or predicted removal ($v_{k,n} = 0$). The binary labels are then randomly and iteratively switched until $P(V|M)$ is maximized. Unlike [16], the significance measure of a coefficient, M, is not merely its Hölder exponent, but evaluated by the magnitude of the coefficients as well as its Hölder approximation, i.e., $f_{M|V}(m_{k,n}|v_{k,n}) = f_{\Lambda|V}(\lambda_{k,n}|v_{k,n})f_{R|V}(\rho_{k,n}|v_{k,n})$. Thus, a joint measure of coefficient significance is developed based on both the Hölder exponent approximation and the magnitude of the wavelet coefficient. As in [16], the coefficients are modified by $\lambda_{k,n}^{\text{new}} = \lambda_{k,n}P(v_{k,n} = 1|M)$.

Although both algorithms in [16] and [21] show promising results in denoised image quality, the iterative procedure necessary to maximize the probability $P(V|M)$ adds computational complexity making the processing times of the algorithms impractical for most image and video processing applications. Also, the Markov random field (MRF) model used in the calculation of $P(V|M)$ is not appropriate for analysis of wavelet coefficients because it ignores the influence of nonneighboring coefficients. The MRF model is strictly used for simplicity and conceptual ease [16].

From the review of the literature, one can see that image denoising remains to be an active and challenging topic of research. The major challenge lies in the fact that one does not know what the original signal is for a corrupted image. The performance of a method can only be statistically estimated or practically measured by comparing the denoised image with its known origin. In this paper, we propose a new denoising approach which delivers better performance than the best results of the previous works of [16] and [21]. The approach consists of two components. The first component is the selective wavelet shrinkage method, and the second is a two-threshold criterion in determining the significance of a wavelet coefficient for its selection or rejection.

The first threshold of the two-threshold selection method determines the importance of a wavelet coefficient by its magnitude, and the second determines its importance by its spatial support of neighboring wavelet coefficients which as a group represent a feature residing in the original image; thus, a feature-based selective wavelet shrinkage method. Furthermore, the method is based upon minimizing the error between the wavelet coefficients of the denoised image and the wavelet coefficients of an optimally denoised image produced by a method using supplemental information. The supplemental information provided produces a denoised image that is far superior than any method which does not utilize supplemental information. Thus, the image produced by the method utilizing supplemental information is referred to as an optimally denoised image. Using several test cases, the two threshold values which produce the minimum difference between the wavelet coefficients of the denoised image and the wavelet coefficients of the optimally denoised image are chosen as the threshold values for the general case.

The two-threshold coefficient selection method results in a denoising algorithm which gives improved results upon those provided by [16] and [21], but without the computational complexity. The two-threshold requirement investigates the regularities of wavelet coefficients both spatially and across scales for predictive coefficient selection, providing selective wavelet shrinkage to nondecimated wavelet subbands.

Following the Introduction, Section II gives theory on the two-dimensional (2-D) nondecimated wavelet analysis and synthesis filters. Section III gives the theoretical base of the proposed method. Section IV then describes the coefficient selection process prior to selective wavelet shrinkage. Section V gives testing results for parameter selection. Section VI gives the estimation algorithms for proper parameter selection, and Section VII gives the results. Section VIII concludes the paper.

II. Two-Dimensional Nondecimated Wavelet Analysis and Synthesis

To facilitate the discussion of the proposed method, nondecimated wavelet filterbank theory is presented. In certain applications, such as signal denoising, it is not desirable to downsample wavelet coefficients after decomposition, as in the tradition wavelet filterbank. The spatial resolution of the coefficients is degraded due to downsampling. Therefore, for

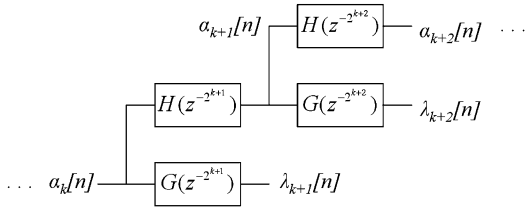


Fig. 1. Nondecimated wavelet decomposition.

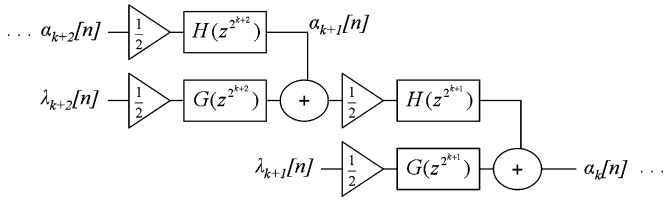


Fig. 2. Nondecimated wavelet synthesis.

the nondecimated case, each subband contains the same number of coefficients as the original signal.

Let $a_k[n]$ and $d_k[n]$ be scaling and wavelet coefficients, respectively, of scale k and position n. Also let $h[\cdot]$ and $g[\cdot]$ be the filter coefficients corresponding to the low-pass and high-pass filter, respectively, of the wavelet transform.

Thus

$$\begin{aligned}\alpha_k[2^{k+1}n] &= a_k[n]\\ \lambda_k[2^{k+1}n] &= d_k[n]\end{aligned} \tag{1}$$

where $\alpha_k[\cdot]$ are the nondecimated scaling function coefficients, and $\lambda_k[\cdot]$ are the nondecimated wavelet coefficients. Equation (1) is substituted into the scaling analysis filterbank equation to find the nondecimated filterbank equation

$$\begin{aligned}a_{k_{+1}}[n] &= \sum_m h[m]a_k[m-2n]\\ \alpha_{k_{+1}}[2^{k+2}n] &= \sum_m h[m]\alpha_k[2^{k+1}(m-2n)]\\ \alpha_{k_{+1}}[n] &= \sum_m h[m]\alpha_k[2^{k+1}m-n].\end{aligned} \tag{2}$$

The 2^{k+1} scalar introduced into (2) is equivalent to upsampling $h[\cdot]$ by 2^{k+1} prior to its convolution with $\alpha_k[\cdot]$. Similarly, (1) is substituted into the wavelet analysis filterbank equation to obtain

$$\lambda_{k_{+1}}[n] = \sum_m g[m]\alpha_k[2^{k+1}m-n]. \tag{3}$$

Fig. 1 gives a block diagram of the nondecimated wavelet decomposition.

The synthesis of the nondecimated wavelet transform also differs from the downsampled case. From the wavelet synthesis filterbank equation, we obtain

$$\begin{aligned}a_k[2n] = &\sum_m h[2(n-m)]a_{k_{+1}}[m]\\ &+\sum_m g[2(n-m)]d_{k_{+1}}[m].\end{aligned} \tag{4}$$

Substituting $(p = n-m)$, we obtain

$$a_k[2n] = \sum_p h[2p]a_{k_{+1}}[n-p] + \sum_p g[2p]d_{k_{+1}}[n-p]. \tag{5}$$

Substituting (1) into (5)

$$\begin{aligned}\alpha_k[2^{k+2}n] = &\sum_p h[2p]\alpha_{k_{+1}}[2^{k+2}(n-p)]\\ &+\sum_p g[2p]\lambda_{k_{+1}}[2^{k+2}(n-p)]\end{aligned} \tag{6}$$

and

$$\begin{aligned}\alpha_k[n] = &\sum_p h[2p]\alpha_{k_{+1}}[n-2^{k+2}p]\\ &+\sum_p g[2p]\lambda_{k_{+1}}[n-2^{k+2}p].\end{aligned} \tag{7}$$

Looking at (7), samples are being thrown away by downsampling $\alpha_{k_{+1}}[\cdot]$ and $\lambda_{k_{+1}}[\cdot]$ by 2 prior to convolution. Because the downsampling in the analysis filters is eliminated, a downsample by 2 is shown in the synthesis equation, (7). If a downsample by 2 is not performed, i.e., $(m = 2p)$, then we must divide by 2 to provide power equality. That is

$$\begin{aligned}\alpha_k[n] = &\frac{1}{2}\sum_m h[m]\alpha_{k_{+1}}[n-2^{k+1}m]\\ &+\frac{1}{2}\sum_m g[m]\lambda_{k_{+1}}[n-2^{k+1}m].\end{aligned} \tag{8}$$

Fig. 2 gives a block diagram of the nondecimated wavelet transform synthesis.

The above analysis is expanded to the 2-D case. For a 2-D discrete signal, $f(\cdot)$

$$\begin{aligned}\alpha_{ll,k_{+1}}[x,y] &= \sum_{n,m} h[n]h[m]\alpha_{ll,k}[2^{k+1}m-x, 2^{k+1}n-y]\\ \lambda_{hl,k_{+1}}[x,y] &= \sum_{n,m} h[n]g[m]\alpha_{ll,k}[2^{k+1}m-x, 2^{k+1}n-y]\\ \lambda_{lh,k_{+1}}[x,y] &= \sum_{n,m} g[n]h[m]\alpha_{ll,k}[2^{k+1}m-x, 2^{k+1}n-y]\\ \lambda_{hh,k_{+1}}[x,y] &= \sum_{n,m} g[n]g[m]\\ &\quad\times \alpha_{ll,k}[2^{k+1}m-x, 2^{k+1}n-y]\end{aligned} \tag{9}$$

where

$$\alpha_{ll,-1}[x,y] = f(x,y). \tag{10}$$

The four coefficient sets given in (9) is referred to as the low–low band $\alpha_{ll,k_{+1}}[\cdot]$, the high–low band $\lambda_{hl,k_{+1}}[\cdot]$, the low–high band $\lambda_{lh,k_{+1}}[\cdot]$, and the high–high band $\lambda_{hh,k_{+1}}[\cdot]$. The subbands are named due to the order in which the scaling and/or the wavelet filters process the scaling function coefficients.

For the synthesis of $f(\cdot)$

$$\begin{aligned}\alpha_{ll,k}[x,y] = &\frac{1}{4}\sum_{m,n} h[m]h[n] \\ &\alpha_{ll,k_{+1}} \times [x-2^{k+1}m, y-2^{k+1}n] \\ &+\frac{1}{4}\sum_{m,n} h[m]g[n] \\ &\lambda_{hl,k_{+1}} \times [x-2^{k+1}m, y-2^{k+1}n] \\ &+\frac{1}{4}\sum_{m,n} g[m]h[n] \\ &\lambda_{lh,k_{+1}} \times [x-2^{k+1}m, y-2^{k+1}n] \\ &+\frac{1}{4}\sum_{m,n} g[m]g[n] \\ &\lambda_{hh,k_{+1}} \times [x-2^{k+1}m, y-2^{k+1}n]. \quad (11)\end{aligned}$$

Equation (9) is recursively computed to produce several levels of wavelet coefficients, and the reconstruction of the 2-D signal, $f(\cdot)$ is accomplished by the recursive computation of (11).

The nondecimated wavelet transform has a number of advantages in signal denoising over the traditional decimated case. One, each subband in the wavelet decomposition is equal in size, and, thus, it is more straightforward to find the spatial relationships between subbands. Two, the spatial resolution of each of the subbands is preserved by eliminating the downsample by two. Because of the elimination of the downsampler, information contained in the wavelet coefficients is redundant, and this redundancy is exploited to determine the coefficients comprised of noise and the coefficients comprised of feature information contained in the original image.

III. Theoretical Base of the Proposed Method

As mentioned in the introduction, wavelet shrinkage methods reduce the magnitude of wavelet coefficients of corrupted images by a parameter $c[\cdot]$ between 0 and 1

$$L_{\cdot,k}[x,y] = c[x,y]\widetilde{\lambda}_{\cdot,k}[x,y] \quad (12)$$

where $L_{\cdot,k}[x,y]$ is the modified wavelet coefficient of scale k and spatial location (x,y), $\cdot \in \{lh, hl, hh\}$ and $\widetilde{\lambda}_{\cdot,k}[x,y]$ is the wavelet coefficient of the corrupted image. In [16] and [21], $c[\cdot]$ is selected to be the marginal probability $P(v=1|M)$. The selective wavelet shrinkage method uses binary shrinkage, $c[\cdot]=0$ or $c[\cdot]=1$. That is, $\widetilde{\lambda}_{\cdot,k}[\cdot]$ is either selected or rejected according to the significance of the wavelet coefficients. Unfortunately, no theory has been provided by any previous works regarding the performance of the two methods. It is obvious that one can only determine the performance statistically because the noise which corrupts the image is assumed to be random with a normal distribution.

The goal of the proposed method, as well as the methods of [16] and [21], is to minimize the difference between the original image and the denoised image. Similarly, a quality metric known as PSNR is used for comparison purposes and is given by

$$\mathrm{PSNR} = 20\log_{10}\left(\frac{255}{\sqrt{\mathrm{mse}}}\right) \quad (13)$$

where

$$\mathrm{mse} = \frac{1}{W_f H_f}\sum_x\sum_y \left(\hat{f}(x,y) - f(x,y)\right)^2. \quad (14)$$

mse is the mean-squared error (MSE) between the original image $f(\cdot)$ and the denoised image $\hat{f}(\cdot)$, and W_f and H_f are the width and height of the image, respectively. In the above equation, mse represents the power of the difference between the original and denoised images. According to Parseval's Theorem [1], minimizing (14) is equivalent to minimizing

$$\mathrm{mse}_\lambda = \frac{1}{M}\sum (c[x,y]\widetilde{\lambda}_{\cdot,k}[x,y] - \lambda_{\cdot,k}[x,y])^2 \quad (15)$$

where $\lambda_{\cdot,k}[x,y]$ is the wavelet coefficient of the original image at location (x,y) and scale k, and $\widetilde{\lambda}_{\cdot,k}[x,y] = \lambda_{\cdot,k}[x,y] + N(x,y)$, where $N(x,y)$ is the noise represented in the wavelet domain. M is the number of wavelet coefficients. Since $\widetilde{\lambda}_{\cdot,k}[x,y]$ is a random variable, and $\lambda_{\cdot,k}[x,y]$ is unknown, we can only minimize the expectation of mse_λ, i.e., $\min[E(\mathrm{mse}_\lambda)]$. The goal is to find $c[\cdot]$ such that $E(\mathrm{mse}_\lambda)$ is minimized. Taking the derivative of $E(\mathrm{mse}_\lambda)$ with respect to $c[\cdot]$, one obtains

$$\begin{aligned}&\frac{\partial_{E(\mathrm{mse}_\lambda)}}{\partial_{c[x,y]}} \\ &= E\left(-2\lambda^2_{\cdot,k}[x,y] + 2c[x,y]\left(\lambda^2_{\cdot,k}[x,y] + N^2(x,y)\right)\right) \quad (16)\end{aligned}$$

using the assumption that $\lambda_{\cdot,k}[\cdot]$ and $N(\cdot)$ are independent, and $N(\cdot)$ is zero mean. To minimize $E(\mathrm{mse}_\lambda)$, (16) is set equal to zero, which leads to

$$c[x,y] = \frac{1}{1+\frac{N^2(x,y)}{\lambda^2_{\cdot,k}[x,y]}}. \quad (17)$$

Equation (17) means that if one wants to maximize the PSNR of the denoised image, $c[\cdot]$ has to be selected individually according to the ratio between $N^2(\cdot)$ and $\lambda^2_{\cdot,k}[\cdot]$. Note that (17) is similar to the classical Weiner filter which minimizes the difference between the power spectra of the original and the corrupted signals [9]. The difference here is that the classical Weiner is derived through the Fourier transform over the correlation between the original signal and noise, which leads to the ratio between the power spectra of the original signal and the noise. Here the energy of the difference is represented by (15) using the individual wavelet coefficients due to Parseval's Theorem applied in the wavelet domain.

Unfortunately, $N(\cdot)$ is a random variable and $\lambda_{\cdot,k}[\cdot]$ is not known once corrupted. Consequently, $N^2(\cdot)/\lambda^2_{\cdot,k}[\cdot]$ can only be

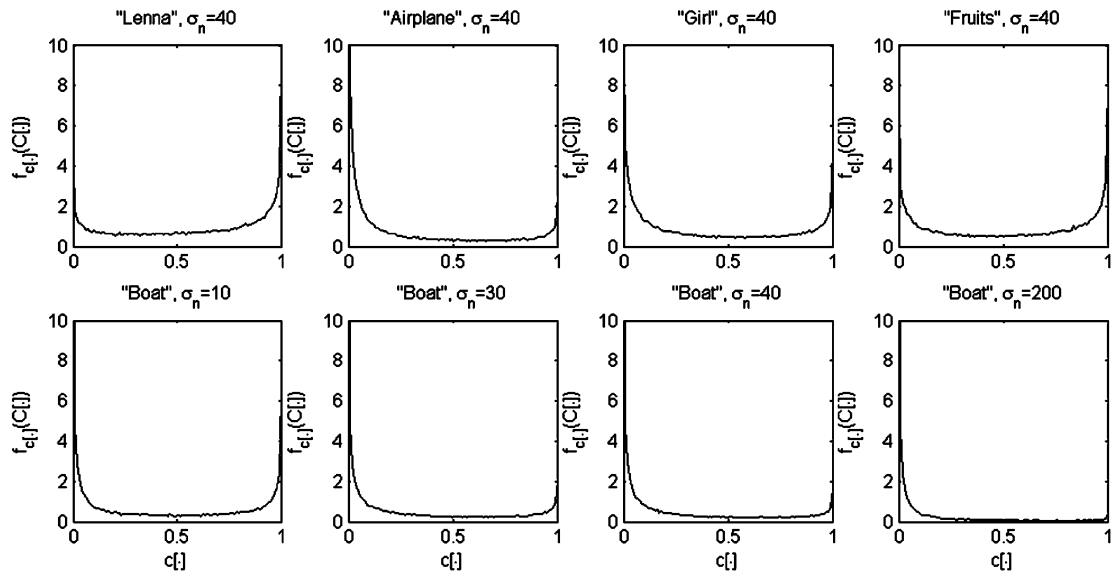


Fig. 3. Probability density function of $c[\cdot]$ generated empirically from (top) four test images with an identical level of random noise and (bottom) one test image with four different levels of random noise.

estimated based on corrupted wavelet coefficients, $\widetilde{\lambda}_{\cdot,k}[\cdot]$, and the neighboring wavelet coefficients which support $\widetilde{\lambda}_{\cdot,k}[\cdot]$ at different levels of scale. Statistically we are to find the best estimate of $N^2(\cdot)/\lambda^2_{\cdot,k}[\cdot]$ given the values of those wavelet coefficients, i.e.,

$$\frac{\hat{N}^2(x,y)}{\lambda^2_{\cdot,k}[x,y]} = f\left(\widetilde{\lambda}_{\cdot,k}[x,y], S_{\cdot,k}[x,y], \widetilde{\lambda}_{\cdot,k+1}[x,y]\right) \qquad (18)$$

where $S_{\cdot,k}[x,y]$ is number of the wavelet coefficients which support $\widetilde{\lambda}_{\cdot,k}[\cdot]$ at scale level k and $\widetilde{\lambda}_{\cdot,k+1}[\cdot]$ is the wavelet coefficient at level $k+1$. Unfortunately, there is no explicit form for the function given in (18). We use two steps to estimate $N^2(x,y)/\lambda^2_{\cdot,k}[\cdot]$. The first step is to judge if $\widetilde{\lambda}_{\cdot,k}[\cdot]$ is dominated by $\lambda_{\cdot,k}[\cdot]$ or by noise. Once the dominating factor is determined, we estimate $N^2(x,y)/\lambda^2_{\cdot,k}[\cdot]$ and consequently $c[\cdot]$.

As mentioned earlier, $\widetilde{\lambda}_{\cdot,k}[\cdot]$ must be greater than a threshold to be valid, and secondly a valid coefficient is selected if it is supported by its neighbor wavelet coefficients at the same level. Furthermore, if the higher level wavelet coefficient covering the same spatial position, i.e., $\widetilde{\lambda}_{\cdot,k+1}[\cdot]$ is valid, the current coefficient should be selected too. The final criterion is based on an observation by previous works on wavelet-based image compression. That is, when a parent coefficient is significant and selected at level $k+1$, its child coefficients are likely to be significant and should be selected [23].

Determining if a wavelet coefficient is dominated by $\lambda_{\cdot,k}[\cdot]$ plays a very important part in the current approach because the binary mechanism simply keeps the coefficient without any further computation. In the following we discuss why the binary mechanism is better that the continuous mechanism once a wavelet coefficient is judged to primarily represent feature or noise.

In statistics, to obtain the value of $N^2(\cdot)/\lambda^2_{\cdot,k}[\cdot]$ using samples of random variables is called a point estimate. It is not possible to tell how close a point estimate is to its true value. Statistically, merit of a point estimator can only be measured by the probability that the true value of a parameter falls in an interval [2] which is called the interval estimate. A better point estimate produces narrower interval with a higher confidence level. In the following we show that the binary shrinkage approach produces a better point estimate than the continuous approach as used in [16], [21].

To find the performance of a point estimate of $N^2(\cdot)/\lambda^2_{\cdot,k}[\cdot]$, it is necessary to know the distribution of $N^2(\cdot)/\lambda^2_{\cdot,k}[\cdot]$. $N[\cdot]$ is assumed to have a normal distribution while $\lambda_{\cdot,k}[\cdot]$ is assumed to have an exponential distribution [21]. One can see easily that $N^2(\cdot)/\lambda^2_{\cdot,k}[\cdot]$ is distributed between 0 and ∞, although it is not possible to formulate the distribution of $N^2(x,y)/\lambda^2_{\cdot,k}[\cdot]$ exactly. From (17) it is realized that the distribution of $c[\cdot]$ exists from 0 to 1.

We use an empirical approach to find the distribution of $c[\cdot]$ using some images which are artificially corrupted by white noise. Fig. 3 shows such histograms derived from several test images and a noise function with a number of variances. Values of $c[\cdot]$ are given in Fig. 3, and the images used in the generations of $c[\cdot]$ are given in Fig. 7. The obtained histograms serve as the estimated probability density function of $c[\cdot]$. One can find that the distribution has a bowl shape. That is, $c[\cdot]$ has a higher probability to be either 0 or 1 than any other value. Using the performance of the point estimate, one can find if $c[\cdot]$ is greater than 0.5 or $\lambda_{\cdot,k}[\cdot]$ is judged to dominate in $\widetilde{\lambda}_{\cdot,k+1}[\cdot]$ based on the samples of random variables which are shown in (18), $c[\cdot]=1$ is a better point estimate than any other estimate. This is because a narrow interval which is between 1 and $1-\alpha$ where α is small number will have a higher confidence level than any other interval of same length except the one between 0 and α. The latter is not possible because $\lambda_{\cdot,k}[\cdot]$ already dominates $\widetilde{\lambda}_{\cdot,k}[\cdot]$. On the other hand, if $N[\cdot]$ dominates $\widetilde{\lambda}_{\cdot,k}[\cdot]$, 0 is the best point estimate of $c[\cdot]$.

From the above discussion, one can see that 0 or 1 is the best estimate statistically using the theory of the point estimate. The continuous approaches using $P(v=1|M)$ as an estimate of $c[\cdot]$

is not able to give a narrow interval with a specific confidence level since $c[\cdot] = P(v = 1|M)$ could be any value between 0 and 1.

IV. Retention of Feature-Supporting Wavelet Coefficients

In the previous section, we have shown that the selective wavelet shrinkage method could have a better performance than the probabilistic approach if the significance of the wavelet coefficient can be determined. In this section, we discuss how the significance is determined in our approach.

One of the many advantages of the wavelet transform over other mathematical transformations is the retention of the spatial relationship between pixels in the original image by the coefficients in the wavelet domain. These spatial relationships represent features of the image and should be retained as much as possible during denoising. In general, images are comprised of regular features, and the resulting wavelet transform of an image generates few large, spatially contiguous coefficients, which are representative of the features given in the original image. We refer to the spatial contiguity of the wavelet coefficients as *spatial regularity*.

The concept of spatial regularity has the similar function as that of signal regularity in previous denoising approaches for selecting the wavelet coefficients. The key difference is that spatial correlation of the features is represented by connectivity of wavelet coefficients rather than statistical models such as Markov random sequences [16], [21] or (Hölder) exponents [16], [17], [21] in previous methods. These models are often computationally complicated and still do not reflect the geometry of the features explicitly. As a result, the current method has a better performance even with a much simpler computation.

Because of spatial regularity, the resulting subbands of the wavelet transform do not generally contain isolated coefficients. This regularity can aid in deciding which coefficients should be selected for reconstruction, and which should be thrown away for maximum reconstructed image quality. The proposed coefficient selection method in which spatial regularity is exploited is shown as follows.

Assume that an image is corrupted with additive noise, i.e.,

$$\widetilde{f}(x,y) = f(x,y) + \eta(x,y) \tag{19}$$

where $f(x,y)$ is the noiseless 2-D signal, $\eta(x,y)$ is a random noise function, and $\widetilde{f}(x,y)$ is the corrupted signal.

The first step for selecting the wavelet coefficient is to form a preliminary binary label for each coefficient, which collectively form a binary map. The binary map is then used to determine whether or not a particular wavelet coefficient is included in a regular spatial feature. The wavelet transform of $\widetilde{f}(x,y)$ generates coefficients, $\widetilde{\lambda}_{\cdot,k}[\cdot]$, from (9) and (10). $\widetilde{\lambda}_{\cdot,k}[\cdot]$ is used to create the preliminary binary map, $I_{\cdot,k}[\cdot]$

$$I_{\cdot,k}[x,y] = \begin{cases} 1, & \text{when } |\widetilde{\lambda}_{\cdot,k}[x,y]| > \tau \\ 0, & \text{else} \end{cases} \tag{20}$$

where τ is a threshold for selecting *valid coefficients* in the construction of the binary coefficient map. A valid coefficient is defined as a coefficient, $\widetilde{\lambda}_{\cdot,k}[x,y]$, which results in $I_{\cdot,k}[x,y] = 1$; hence, the coefficient has been selected due to its magnitude. After coefficients are selected by magnitude, spatial regularity is used to further examine the role of the valid coefficient: whether it is isolated noise or part of a spatial feature. The number of supporting binary values around a particular nonzero value $I_{\cdot,k}[x,y]$ is used to make the judgement. The support value, $S_{\cdot,k}[x,y]$, is the sum of all $I_{\cdot,k}[\cdot]$ which support the current binary value $I_{\cdot,k}[x,y]$; that is, the total number of all valid coefficients which are spatially connected to $I_{\cdot,k}[x,y]$.

A coefficient is spatially connected to another if there exists a continuous path of valid coefficients between the two. Fig. 4 gives a generic coefficient map. The valid coefficients are highlighted in gray. From Fig. 4 it can be shown that coefficients A, B, C, and H do not support any other valid coefficients in the coefficient map. However, coefficients D and F support each other, coefficients E and G support each other, and N and O support each other. Also, coefficients I, J, K, L, M, P, Q, and R all support one another. Fig. 5 gives the value of $S_{\cdot,k}[x,y]$ for each of the valid coefficients given in Fig. 4. A method of computing $S_{\cdot,k}[x,y]$ is given in Appendix. $S_{\cdot,k}[\cdot]$ is used to refine the original binary map $I_{\cdot,k}[\cdot]$ by (21), shown at the bottom of the page, where $J_{\cdot,k}[\cdot]$ is the refined binary map, and s is the necessary number of support coefficients for selection. $J_{\cdot,\cdot}[\cdot]$ is calculated recursively, starting from the highest multiresolution level, and progressing downward.

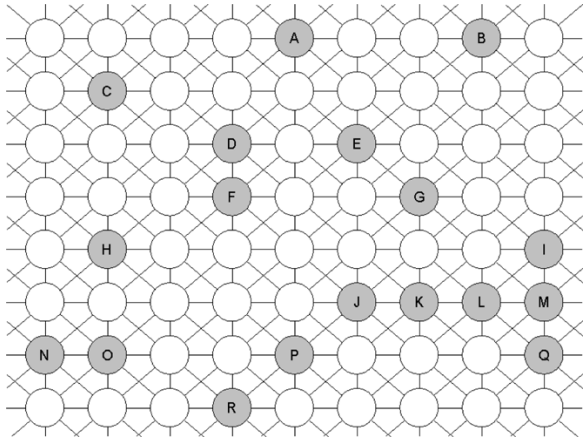


Fig. 4. Generic coefficient array.

$$J_{\cdot,k}[x,y] = \begin{cases} 1, & \text{when } S_{\cdot,k}[x,y] > s, \text{ or } J_{\cdot,k+1}[x,y] I_{\cdot,k}[x,y] = 1 \\ 0, & \text{else} \end{cases} \tag{21}$$

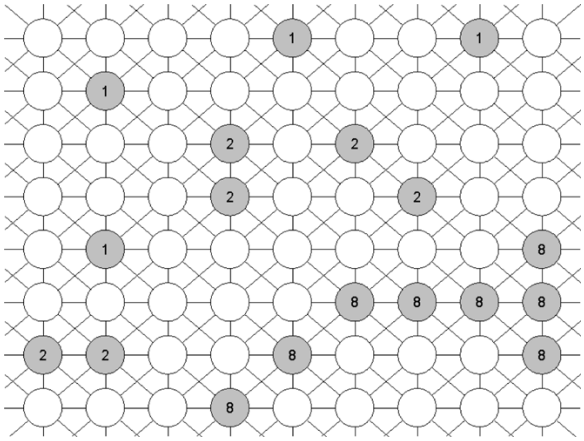


Fig. 5. Generic coefficient array, with corresponding $S_{\cdot,k}$ values.

Equation (21) is equal to one when there exists enough wavelet coefficients of large magnitude around the current coefficient. However, it also is equal to one when the magnitude of the coefficient is effectively large $(I_{\cdot,k}[\cdot] = 1)$, but not locally supported $(J_{\cdot,k}[\cdot] = 0)$ only if the coefficient of the larger scale is large and locally supported $(J_{\cdot,k_{+1}}[\cdot] = 1)$. The decision to use this criterion is in the somewhat rare case when a useful coefficient is not locally supported. In the general case, wavelet coefficients of images are clustered together, but rarely are they isolated. In [17], wavelet coefficients are modified *only* by their evolution across scales. Regular signal features contain wavelet coefficients which increase with increasing scale. Thus, if there exists a useful coefficient which is isolated in an image, it is reasonable that a coefficient in the same spatial location of an increase in scale will be sufficiently large and spatially supported. Thus, the coefficient selection method provided by (22) selects coefficients which are sufficiently large and locally supported as well as isolated coefficients which are sufficiently large and supported by scale.

This type of scale selection is consistent with the findings of Said and Pearlman [23], who developed an image codec based on a "spatial self symmetry" between differing scales in wavelet transformed images. They discovered that most image energy is concentrated in the low-frequency subbands of the wavelet transform. And because of the self-symmetry properties of wavelet transformed images, if a coefficient value is insignificant (i.e., of small value or zero), then it can be assumed that the coefficients of higher spatial frequency and same spatial location will be insignificant. In our application, however, we are looking for significance rather than insignificance, so we look to the significance of lower frequency coefficients to determine significance of the current coefficient. In this way, the preliminary binary map is refined by both spatial and scalar support, given by (21).

The final coefficients retained for reconstruction are given by

$$L_{\cdot,k}[x,y] = \begin{cases} \widetilde{\lambda}_{\cdot,k}[x,y], & \text{when } J_{\cdot,k}[x,y] = 1 \\ 0, & \text{else.} \end{cases} \tag{22}$$

The denoised image is reconstructed using the supported coefficients, $L_{,k}[x,y]$ in the synthesis equation given in (11). Thus

$$\begin{aligned} \hat{\alpha}_{ll,k}[x,y] = \frac{1}{4}\sum_{m,n} & h[m]h[n] \\ & \hat{\alpha}_{ll,k_{+1}} \times [x - 2^{k+1}m, y - 2^{k+1}n] \\ & + \frac{1}{4}\sum_{m,n} h[m]g[n] \\ & L_{hl,k_{+1}} \times [x - 2^{k+1}m, y - 2^{k+1}n] \\ & + \frac{1}{4}\sum_{m,n} g[m]h[n] \\ & L_{lh,k_{+1}} \times [x - 2^{k+1}m, y - 2^{k+1}n] \\ & + \frac{1}{4}\sum_{m,n} g[m]g[n] \\ & L_{hh,k_{+1}} \times [x - 2^{k+1}m, y - 2^{k+1}n]. \end{aligned} \tag{23}$$

Equation (23) is calculated recursively producing scaling coefficients of finer resolution until $k = -1$. The denoised image $\hat{f}(\cdot)$ is then given by

$$\hat{f}(x,y) = \hat{\alpha}_{ll,-1}[x,y]. \tag{24}$$

$\hat{\alpha}_{ll,k}[\cdot]$ are the reconstructed scaling function coefficients.

In general, natural and synthetic imagery can be compactly represented in few wavelet coefficients of large magnitude. These coefficients are in general spatially clustered. Thus, it is useful to obtain selection methods based on magnitude and spatial regularity to distinguish between useful coefficients which are representative of the image and useless coefficients representative of noise. The two-threshold criteria for the rejection of noisy wavelet coefficients is a computationally simple, noniterative test for magnitude and spatial regularity which can effectively distinguish between useful and useless coefficients.

V. Selection of Threshold τ and Support s

The selection of threshold τ and support s is a key component of the denoising algorithm. Unfortunately, the two parameters cannot be easily determined for a given corrupted image because there is no information about the decomposition between the original signal and the noise. We derive τ and s using a set of test images which serve as training samples. These training samples are artificially corrupted by noise. The noise is then removed by a series of τ and s. The set of τ and s which generates the best results is selected for noise removing in general. This approach has its root in an idea called oracle [4] which is described below.

An oracle is an entity which provides extra information to aid in the denoising process. The extra information provided by the oracle is undoubtedly beneficial in providing substantially better denoising results than methods which are not furnished supplemental information. Thus, the coefficient selection method which uses the oracle's information is referred to as the optimal denoising method. By the optimal denoising method, the threshold and support can be selected using test images of which both original image and noise are known. The threshold

Fig. 6. Optimal denoising method applied to noisy "Lena" image. Left: Corrupted image $\widetilde{f}(x,y)$, $\sigma_n = 50$, $\mathrm{PSNR} = 14.16$ dB. Right: Optimally denoised image $\hat{f}^{\mathrm{opt}}(x,y)$, $\mathrm{PSNR} = 27.72$ dB.

and support can then be selected accordingly for any corrupted images without supplemental information.

An optimal coefficient selection process has been defined based on the original (noiseless) image. The optimal binary map $J^{\mathrm{opt}}_{\cdot,k}[\cdot]$ is given by

$$J^{\mathrm{opt}}_{\cdot,k}[x,y] = \begin{cases} 1, & \text{when } |\lambda_{\cdot,k}[x,y]| > \sigma_n \\ 0, & \text{else} \end{cases} \tag{25}$$

where $\lambda_{\cdot,k}[\cdot]$ are the wavelet coefficients of the original (noiseless) image $f(\cdot)$, and σ_n is the standard deviation of the noise in the corrupted image $\widetilde{f}(\cdot)$. Thus, the extra information given by the oracle is the noiseless wavelet coefficients, $\lambda_{\cdot,k}[\cdot]$, and the standard deviation of the noise. The coefficients which are used in the reconstruction, $L^{\mathrm{opt}}_{\cdot,k}[\cdot]$, are given by

$$L^{\mathrm{opt}}_{\cdot,k}[x,y] = \begin{cases} \widetilde{\lambda}_{\cdot,k}[x,y], & \text{when } J^{\mathrm{opt}}_{\cdot,k}[x,y] = 1 \\ 0, & \text{else} \end{cases} \tag{26}$$

where $\widetilde{\lambda}_{\cdot,k}[\cdot]$ are the wavelet coefficients of the noisy image. Note that the method using (25) and (26) does not involve the support of the neighboring coefficients, but only the magnitude. The reason again is the supplemental information. Since the wavelet coefficients of the original image are given, everyone of them should be used in the reconstruction of the denoised image regardless of the support by its neighboring coefficients. Only when a coefficient is severely corrupted by the noise, which could happen when the standard deviation of the noise is greater, it is not included. The approach is in agreement with (17) with an advantage that $\lambda_{\cdot,k}[\cdot]$ is known. As a result, the selection and rejection of a wavelet coefficient has a higher confidence level than without knowing the true value of the wavelet coefficient. The coefficient map thus obtained is considered optimal.

The optimal coefficient map is used to create the optimal denoised image which is given by

$$\begin{aligned} \hat{\alpha}^{\mathrm{opt}}_{ll,k}[x,y] = {} & \frac{1}{4}\sum_m\sum_n h[m]h[n] \\ & \hat{\alpha}^{\mathrm{opt}}_{ll,k_{+1}} \times [x - 2^{k+1}m, y - 2^{k+1}n] \\ & + \frac{1}{4}\sum_m\sum_n h[m]g[n] \\ & L^{\mathrm{opt}}_{hl,k_{+1}} \times [x - 2^{k+1}m, y - 2^{k+1}n] \\ & + \frac{1}{4}\sum_m\sum_n g[m]h[n] \\ & L^{\mathrm{opt}}_{lh,k_{+1}} \times [x - 2^{k+1}m, y - 2^{k+1}n] \\ & + \frac{1}{4}\sum_m\sum_n g[m]g[n] \\ & L^{\mathrm{opt}}_{hh,k_{+1}} \times [x - 2^{k+1}m, y - 2^{k+1}n]. \end{aligned} \tag{27}$$

Equation (27) is recursively computed for lesser values of k until the optimal denoised image is achieved, where

$$\hat{f}^{\mathrm{opt}}(x,y) = \hat{\alpha}^{\mathrm{opt}}_{ll,-1}[x,y]. \tag{28}$$

$\hat{\alpha}^{\mathrm{opt}}_{ll,k}[\cdot]$ are the optimal scaling coefficients, and $\hat{f}^{\mathrm{opt}}(\cdot)$ is the optimal denoised image. Fig. 6 gives the denoising results of the optimal denoising method when applied to the "Lena" image corrupted with additive white Gaussian noise (AWGN). As shown in Fig. 6, the optimal denoising method is able to effectively remove the noise from the "Lena" image because of the added information given by the oracle. PSNR is calculated for performance measurement and is given by (13).

It is rather obvious that the optimal coefficient selection process is unattainable when no supplemental information is provided by the oracle for corrupted images. Thus, the optimal

Fig. 7. Test images.

image denoising method is not possible for practical applications. However, the knowledge obtained by the optimal binary map, $J^{\text{opt}}_{\cdot,k}[\cdot]$, is used to compare with the refined coefficient map generated by the two-threshold criteria, $J_{\cdot,k}[\cdot]$, described in Section IV. The coefficient selection method is based on the error between the optimal coefficient subband and the subband generated by the two-threshold criteria. The error is given by

$$\text{Error} = \frac{\sum_{p,k,x,y} \left(J^{\text{opt}}_{p,k}[x,y] \oplus J_{p,k}[x,y]\right)\left(\widetilde{\lambda}_{p,k}[x,y] - \lambda_{p,k}[x,y]\right)^2}{\sum_{p,k,x,y} \widetilde{\lambda}^2_{p,k}[x,y]} \quad (29)$$

where $\oplus$ is the *exclusive OR* operation. Our goal is to minimize the error using a set of *training samples*.

The proposed coefficient selection algorithm starts with a series of test images serving as training samples to derive the functions which determine the optimal set of the values for τ and s as well as the type of wavelet used for denoising. The original data and the statistical distribution of the noise are given for each of the training samples which are corrupted. The optimal set of parameters can then be determined for the training samples using the approach described earlier. The test images given in Fig. 7 are all 256×256 pixels. Starting from the upper-left image and going clockwise, the images are "Lena," "Airplane," "Girl," "Fruits," "Goldhill," "Boat," "Barb," and "Baboon." Each of the images shown in Fig. 7 is well known in the image-processing community and collectively represents as many kinds of images as possible. In this way, the τ and s obtained will likely perform well in most cases.

A test is used to demonstrate the effectiveness of different wavelets in denoising. First, each of the eight test images is corrupted with AWGN at various levels. Next, the 2-D nondecimated wavelet transform, given in Section II, is calculated using several different wavelets. The wavelet coefficients are then hard thresholded using a threshold T ranging from 0–150, and the inverse wavelet transform is applied to the thresholded coefficients. The wavelet which gives the reconstructed images with the highest average PSNR is chosen to be used in the general case.

Several wavelets were used in the testing. However, for simplicity only five are presented. We have chosen the Daubechies wavelets [3] (Daub4 and Daub8) for their smoothness properties, the spline wavelets (first order and quadradic spline) [1] because of their use in the previous works of [16], [17], [21], and the Haar wavelet because of its simplicity and compact support. The results are given in Fig. 8.

After the testing results given in Fig. 8, the Haar wavelet is selected for image denoising

$$h[n] = \begin{cases} \frac{1}{\sqrt{2}}, & n = 0, 1 \\ 0, & \text{else} \end{cases} \quad g[n] = \begin{cases} \frac{-1}{\sqrt{2}}, & n = 0 \\ \frac{1}{\sqrt{2}}, & n = 1 \\ 0, & \text{else.} \end{cases} \quad (30)$$

Testing has shown the Haar wavelet to be the most promising in providing the highest reconstructed image quality. The compact support of the Haar wavelet enables the wavelet coefficients to represent the least number of original pixels in comparison with other types of wavelets. Therefore, when a coefficient is removed because of its insignificance or isolation, the result affects the smallest area of the original image in the reconstruction. That could reduce the impact to the image quality even if a removed coefficient is not only comprised of noise.

The Haar wavelet is used in a nondecimated wavelet decomposition of the original image. Five subband levels are used, i.e., $k = -1$ to 4. The proposed selective wavelet shrinkage algorithm is applied to all wavelet subbands, and the subbands are synthesized by the nondecimated inverse wavelet transform.

Testing for the optimal values of τ and s is accomplished by artificially adding Gaussian noise to each of the eight images, denoising all eight images with a particular τ and s, and recording the average error given by (29). Then, the combination of τ and s which gives the lowest error is the choice for that particular noise level.

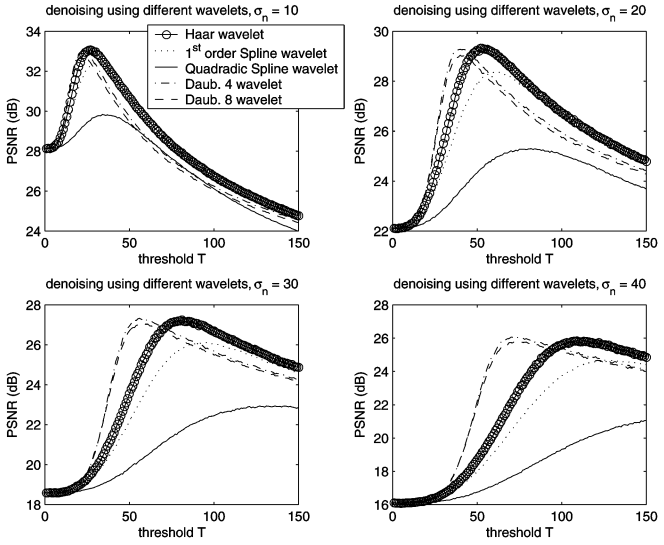


Fig. 8. Average PSNR values using different wavelets.

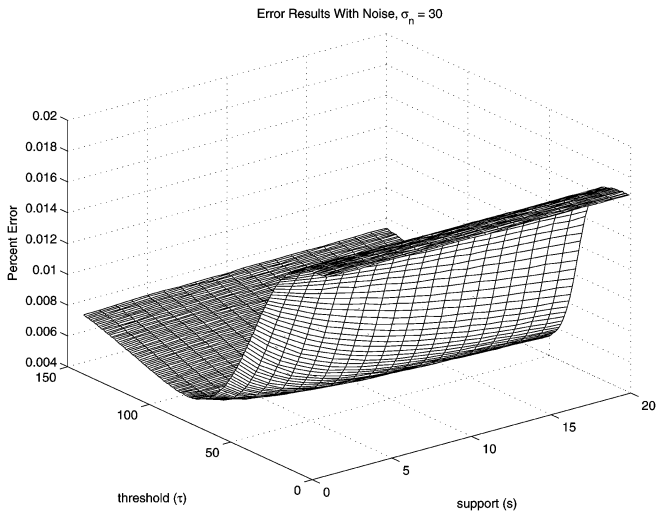


Fig. 9. Error results for test images $\sigma_n = 30$.

The average error is recorded when denoising each of the eight test images given in Fig. 7 using τ ranging from 0–150 and s ranging from 0–20. The proposed algorithm is tested by applying AWGN with a standard deviation (σ_n) of 10, 20, 30, 40, and 50 to each of the test images. The proposed method of selective wavelet shrinkage is applied to the corrupted image, and the resulting error is recorded using (29). The results of the testing in which $\sigma_n = 30$ is given in Fig. 9.

Table I gives the τ and s which provide the lowest average error for each noise level tested. These particular values are referred to as $\tau_m(\cdot)$ and $s_m(\cdot)$. Table I suggests that parameters $\tau_m(\cdot)$ and $s_m(\cdot)$ are functions of the standard deviation of the noise, σ_n.

The two thresholds are obtained by using a number of training samples (eight in the above process). A question natural to ask is how the number of training samples will affect the thresholds? To answer this question, we repeat the above process by using only the top four images in Fig. 7 and receive τ_m and s_m as shown in Table II. One can see that the difference between the results in Tables I and II are relatively small. This proves that

TABLE I
MINIMUM AVERAGE ERROR OF EIGHT TEST IMAGES FOR VARIOUS NOISE LEVELS AND THEIR CORRESPONDING THRESHOLD AND SUPPORT VALUES

Noise Level(σ_n)	10	20	30	40	50
Min. Avg. E_{rror}	5E-4	22E-4	47E-4	77E-4	111E-4
s_m value	6	12	11	12	16
τ_m value	22	44	68	95	115

TABLE II
MINIMUM AVERAGE ERROR OF FOUR TEST IMAGES FOR VARIOUS NOISE LEVELS AND THEIR CORRESPONDING THRESHOLD AND SUPPORT VALUES

Noise Level(σ_n)	10	20	30	40	50
Min. Avg. E_{rror}	5E-4	18E-4	40E-4	68E-4	100E-4
s_m value	5	7	17	19	17
τ_m value	24	45	71	90	116

the proposed method is rather robust to the number of training samples. On the other hand, more training samples do generate better denoising results which will be shown later.

Because $\tau_m(\cdot)$ and $s_m(\cdot)$ generally increase with an increase in additive noise as shown in Table I, both parameters can be modeled as functions of the additive noise, σ_n. Then, knowing the level of noise corruption, the threshold levels which produce the minimum error, given by (29), may be obtained by estimating the $\tau_m(\cdot)$ and $s_m(\cdot)$ functions. The five noise levels provided in the test are used as sampling points for the estimation of the continuous functions $\tau_m(\cdot)$ and $s_m(\cdot)$. With enough sampling points both $\tau_m(\cdot)$ and $s_m(\cdot)$ can be effectively estimated, and the correct τ and s can be calculated to denoise an image with any level of noise corruption, given that the noise level is known.

The estimated functions of the sampled values $\tau_m(\cdot)$ and $s_m(\cdot)$ are referred to as $\widetilde{\tau}_m(\cdot)$ and $\widetilde{s}_m(\cdot)$, respectively. Once the estimated functions are calculated they are used in the general case. Thus, given an image corrupted with noise, it is denoised with no prior knowledge by estimating the level of noise corruption, calculating the proper thresholds using the $\widetilde{\tau}_m(\cdot)$ and $\widetilde{s}_m(\cdot)$ functions and using the calculated threshold levels in the denoising process given in Section IV.

VI. ESTIMATION OF PARAMETER VALUES

It can be shown from the values given in Tables I and II that the parameters $\tau_m(\cdot)$ and $s_m(\cdot)$ are functions of σ_n; therefore, we need to estimate the standard deviation of the noise level, and the functions. These two topics are discussed in this section. Another idea for selecting the two parameters is to use the signal-noise-ratio (SNR) of the image. Unfortunately, the SNR information for a noised image is not given, and very hard to derive if not impossible for reasons mentioned earlier, i.e., one has no idea about the level of the original signal, and has to use an entirely different way to estimate SNR of a corrupted image. On the other hand, there are standard procedures for estimating the standard deviation of the noise level, one of which is shown below.

A. Noise Estimation

The level of noise in a given digital image is unknown and must be estimated from the noisy image data. Several well-

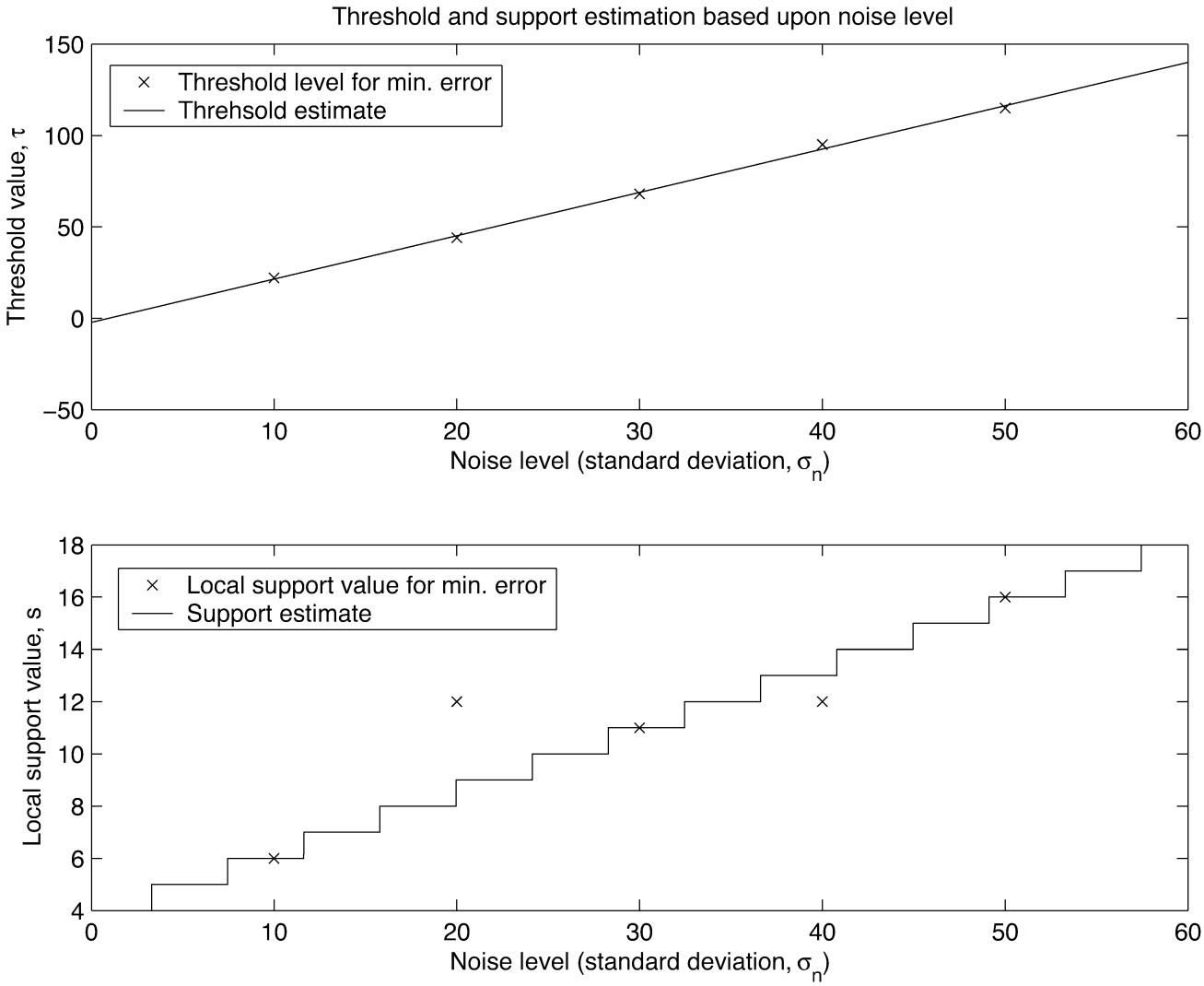


Fig. 10. $\tau_m(\cdot)$, $s_m(\cdot)$ and their corresponding estimates, $\widetilde{\tau_m}(\cdot)$, $\widetilde{s_m}(\cdot)$.

known algorithms have been given in the literature to estimate image noise. From [5], [21] a median value of the $\widetilde{\lambda}_{hh,0}[\cdot]$ subband is used in the estimation process. The median noise estimation method of [21] is used

$$\widetilde{\sigma}_n = \frac{\text{Median}(|\widetilde{\lambda}_{hh,0}[\cdot]|)}{0.6745} \tag{31}$$

where $\widetilde{\lambda}_{hh,0}[\cdot]$ are the noisy wavelet coefficients in the high-high band of the 0th scale. Because the vast majority of useful information in the wavelet domain is confined to few and large coefficients, the median can effectively estimate the level of noise (i.e., the average level of the useless coefficients) without being adversely influenced by useful coefficients.

B. Parameter Estimation

By using the training samples, τ_m and s_m are obtained for a few given standard deviations of the noise. The goal is to obtain τ_m and s_m for any standard deviation. From the training samples, we find that the relationships between τ_m and σ_n and between s_m and σ_n are almost linear as shown in Fig. 10. We, therefore, use a first-order polynomial to represent τ_m and s_m and use the most popular linear minimum mean-square error (LMMSE) method [25] to find the parameters associated with the polynomial. For the first-order polynomial, two parameters a_τ and b_τ are found such that

$$\widetilde{\tau_m}(\sigma_n) = a_\tau \sigma_n + b_\tau. \tag{32}$$

The choice of a_τ and b_τ will minimize the MSE between $\tau_m(\cdot)$ and $\widetilde{\tau_m}(\cdot)$. Similarly, an estimate of s_m, which must be an integer, is found as

$$\widetilde{s_m}(\sigma_n) = \lfloor a_s \sigma_n + b_s \rfloor. \tag{33}$$

Using the test results generated with the eight sample images given in Table I, the parameters which minimize the two MSEs are found to be $a_\tau = 2.37$, $b_\tau = -2.30$, $a_s = 0.24$, and $b_s = 4.21$. In addition, using the test results generated with four sample images given in Table II, the parameters are found to be $a_\tau = 2.29$, $b_\tau = 0.50$, $a_s = 0.34$, and $b_s = 3.81$. Each parameter set is similar to the other, suggesting that the parameter selection method is robust to different sampling image sets.

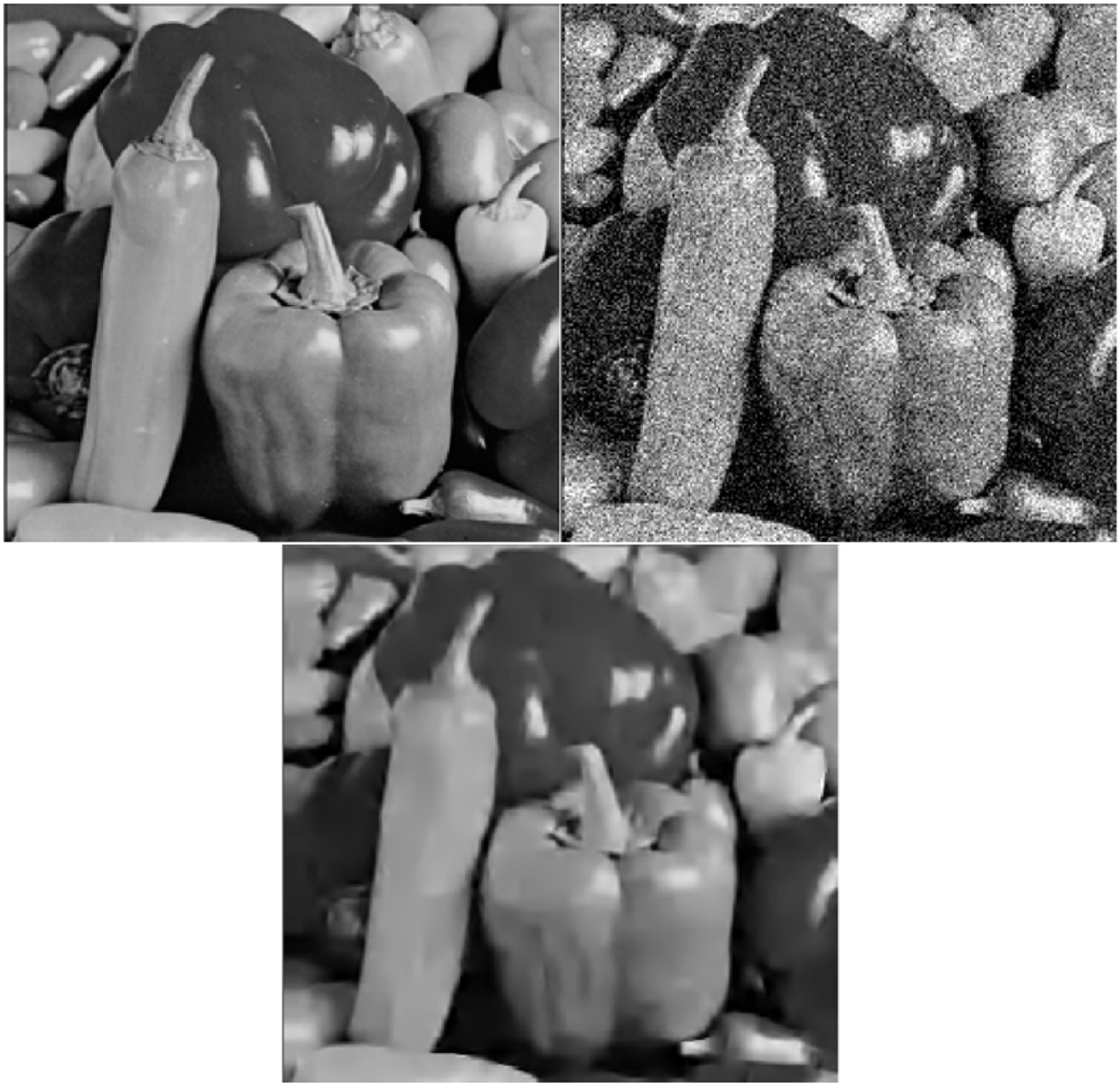


Fig. 11. Results of the proposed image denoising algorithm. Top left: Original "peppers" image. Top right: Corrupted image $\sigma_n = 37.75, \mathrm{PSNR} = 16.60$ dB. Bottom: Denoised image using the proposed method $\mathrm{PSNR} = 27.05$ dB.

The LMMSE estimator is shown to be a good fit into the test data given in Fig. 10. The values of $\tau_m(\cdot)$, and $s_m(\cdot)$ generated by the eight sample images are given, as well as their corresponding LMMSE estimates. Note that the support value s_m must be an integer.

For any corrupted image, the threshold τ and the support value s are determined by using the estimate of the noise given by (31). The two thresholds are given by

$$\begin{aligned} \tau &= a_\tau \widetilde{\sigma_n} + b_\tau \\ s &= \lfloor a_s \widetilde{\sigma_n} + b_s \rfloor. \end{aligned} \tag{34}$$

VII. Experimental Results

The "peppers" and "house" images are used for gauging the performance of the proposed denoising algorithm. These two images have also been used in the results of [16], [17], and [21]. Therefore, the performance of the proposed algorithm is compared with other recent algorithms given in the literature. Both the "peppers" image and "house" image are 256×256 grayscale images which are corrupted with AWGN. The proposed method is used for denoising, and the results are given in Figs. 11 and 12, using the threshold estimates given from the eight test images.

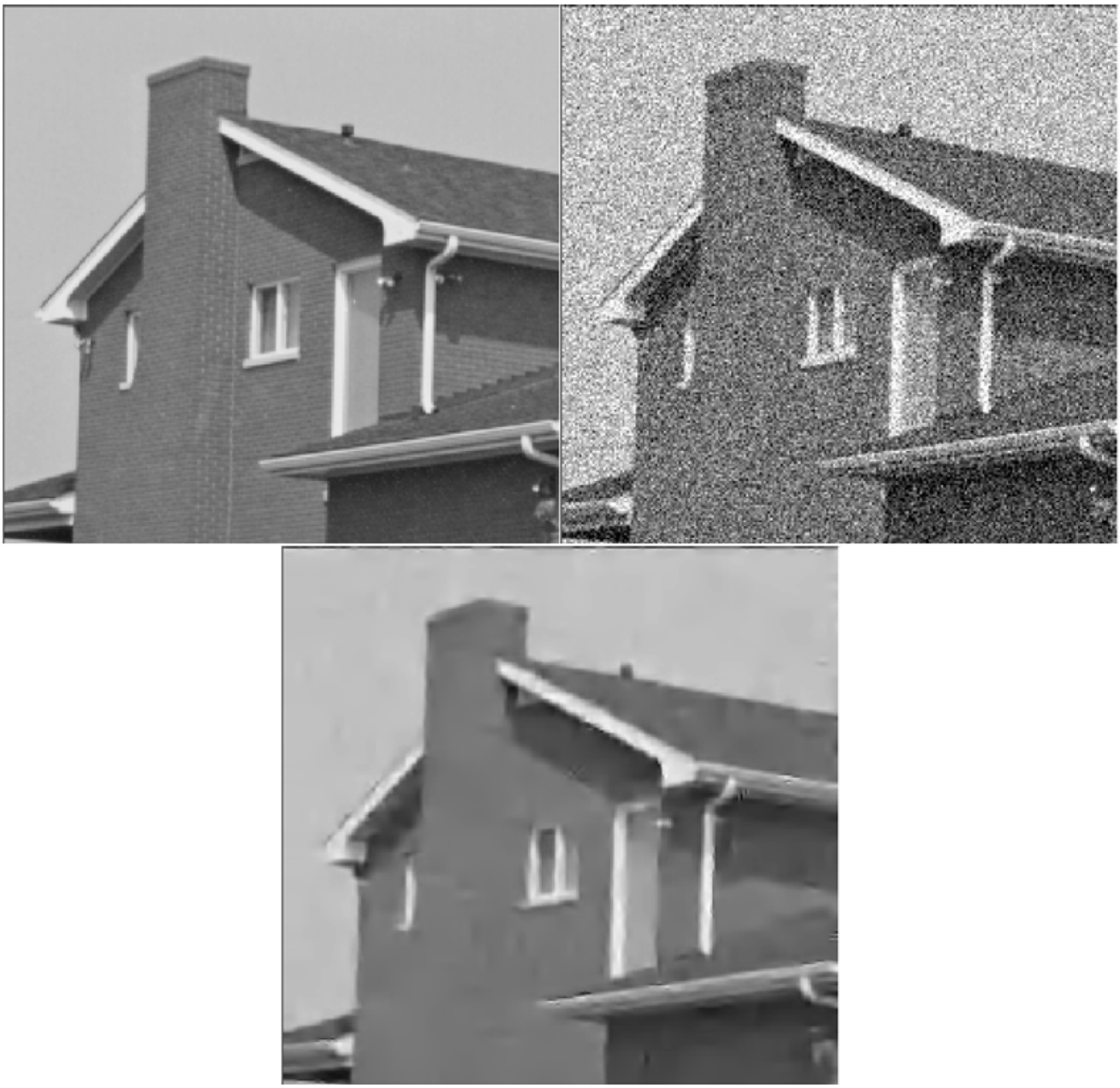


Fig. 12. Results of the proposed image denoising algorithm. Top left; original "house" image. Top right: Corrupted image $\sigma_n = 32.47$, $\mathrm{PSNR} = 17.90$ dB. Bottom: Denoised image using the proposed method $\mathrm{PSNR} = 29.73$ dB.

Table III gives the results of the proposed method, with threshold estimates given from both eight and four test image sets, as well as the results of [16], [17], and [21]. Note that the methods of [16], [17], and [21] all use the quadratic spline wavelet [1], and each of the algorithms' coefficient selection method is based on the Hölder exponent approximation to determine the amount that a particular coefficient contributes to the overall image quality. The proposed algorithm uses the Haar wavelet, given in (30), and the coefficient selection process is based on the two-threshold approach. As shown in Table III, the results of the proposed method are an improvement over other methods described in the literature for both the eight and four training-sample methods. The eight-sample method generates slightly better results which demonstrates that more training samples can denoise different types of images better than few samples as expected. On the other hand, more training samples would consume more training time.

Using the $\tau_m(\cdot)$ and $s_m(\cdot)$ functions determined by the eight training sample images given in Fig. 7, several other images are denoised using the proposed two-threshold selective shrinkage method. A comparison is made between the two-threshold method and the optimal denoising method given in Section V. The results of the additional images are provided in Table IV. All images are 256×256 grayscale and shown in Fig. 13. Clockwise starting from the upper left, the images are

TABLE III
PSNR COMPARISON OF THE PROPOSED METHOD TO OTHER METHODS IN THE LITERATURE (IN DECIBELS)

"Peppers"					
Image Input PSNR	22.6	19.6	16.6	13.6	Average
Proposed Algorithm (8 sample images)	30.90	28.89	27.05	25.28	28.03
Proposed Algorithm (4 sample images)	31.00	28.68	26.86	25.21	27.94
Pizurica 3-band, [21]	30.20	28.60	27.00	25.20	27.75
Pizurica 2-band, [21]	29.90	28.20	26.60	24.90	27.40
Malfait and Roose, [16]	28.60	27.30	26.00	24.60	26.63
Mallat and Hwang, [17]	28.20	27.30	27.10	24.60	26.80
Matlab's Sp. Adaptive Wiener	29.00	27.10	25.30	23.30	26.18
"House"					
Image Input PSNR	23.9	20.9	17.9	14.9	Average
Proposed Algorithm (8 sample images)	33.06	31.61	29.73	28.37	30.69
Proposed Algorithm (4 sample images)	32.91	31.53	29.67	28.11	30.56
Pizurica 3-band, [21]	32.80	31.30	29.80	28.30	30.55
Pizurica 2-band, [21]	32.10	30.50	29.30	28.10	30.00
Malfait and Roose, [16]	32.90	31.30	29.80	28.20	30.55
Mallat and Hwang, [17]	31.30	30.50	29.10	27.10	29.50
Matlab's Sp. Adaptive Wiener	30.30	28.60	26.70	24.90	27.63

Fig. 13. Various Images used for performance testing.

"Lake," "Camera," "Tiffany," and "Elaine." One can see from Table IV that the optimal denoising approach produces better results as expected. This result is consistent with the purpose of the oracle approach and shows that the proposed denoising method is effective by minimizing the difference to the result of the optimal denoising approach.

In addition to improved performance over previous methods, the proposed algorithm is computationally simple to facilitate real-world applications. The proposed algorithm has been computed on older processors for an accurate comparison, and the computation time of the proposed method is over an order of magnitude less than the previous method of highest performance, [21]. Table V gives the computational results of the proposed method as well as the results of [16], and [21]. The proposed algorithm shows a substantial drop in computation time. Both [16] and [21] used iterative computation in the selection of wavelet coefficients for reconstruction which requires unreasonable computation time for certain applications. The current two-threshold technique is a simpler, noniterative coefficient selection method which produces greater performance results.

VIII. CONCLUSION

In this paper, a new selective wavelet shrinkage algorithm for image denoising has been described. We have shown that the selective wavelet shrinkage method which either selects or rejects a wavelet coefficient is statistically better than the probabilistic method because the former can identify a narrow interval for the estimated parameter, which is used to adjust the wavelet coefficient, with a higher confidence level than the latter. For selecting or rejecting a wavelet coefficient, the proposed algorithm uses a two-threshold support criteria which investigates coefficient magnitude, spatial support, and support across scales in the coefficient selection process. In general, images can be accurately represented by a few large wavelet coefficients, and those few coefficients are spatially clustered together. The two-threshold criteria is an efficient and effective way of using the magnitude and spatial regularity of wavelet coefficients to distinguish useful from useless coefficients. Furthermore, the two-threshold criteria is a noniterative solution to selective wavelet shrinkage to provide a computationally simple solution, facilitating real-time image-processing applications.

The values of the two-thresholds are determined by minimizing the error between the coefficients selected by the two-thresholds and the coefficients selected by a denoising method which uses supplemental information provided by an oracle. The supplemental information provided by the oracle is useful in determining the correct coefficients to select, and the denoising performance is substantially greater than methods which do not use the supplemental information. Thus, the method which uses the supplemental information provided by the oracle is referred to as the optimal denoising method.

TABLE IV
RESULTS OF PROPOSED AND OPTIMAL METHODS ON VARIOUS IMAGES

Image Input PSNR	23.0	20.0	17.0	14.0	Average
"Lake"					
Optimal Method	30.25	28.68	26.89	25.42	27.81
Proposed Method (8 sample images)	26.38	24.27	22.61	21.21	23.62
Proposed Method (4 sample images)	26.04	24.13	22.50	21.13	23.45
"Camera"					
Optimal Method	32.62	30.85	29.08	27.51	30.01
Proposed Method (8 sample images)	29.41	27.88	25.71	24.22	26.81
Proposed Method (4 sample images)	29.21	27.69	25.51	24.14	26.64
"Tiffany"					
Optimal Method	31.45	29.88	28.42	26.87	29.16
Proposed Method (8 sample images)	28.24	26.55	24.88	23.54	25.80
Proposed Method (4 sample images)	27.96	26.35	24.98	23.52	25.70
"Elaine"					
Optimal Method	32.20	30.86	29.64	28.27	30.24
Proposed Method (8 sample images)	30.15	28.48	26.93	25.38	27.74
Proposed Method (4 sample images)	29.83	28.47	26.83	25.17	27.58

TABLE V
COMPUTATION TIMES FOR A 256×256 IMAGE, IN SECONDS

Processor	Pentium IV	Pentium III	IBM RS6000/320H
Proposed Algorithm	0.66	1.14	***
Pizurica 3-band, [21]	***	45.00	***
Pizurica 2-band, [21]	***	30.00	***
Malfait and Roose, [16]	***	***	180.00

*** Computation time not evaluated

Therefore, by minimizing the error between the two-threshold method and the optimal denoising method, the two-threshold method can come as close as possible to the performance of the optimal denoising method.

Consequently, the two-threshold method of selective wavelet shrinkage provides an image denoising algorithm which is superior to previous image denoising methods given in the literature both in denoised image quality and computation time. The light computational burden of the proposed denoising method makes it suitable for real-time image-processing applications.

APPENDIX

The computation of $S_{\cdot,k}[x,y]$ is given by the following algorithm:

$\vec{N}() =$
$\{[-1,-1],[-1,0],[-1,1],[0,-1],[0,1],[1,-1],[1,1],\}$
$O[\cdot] = 0,\ \ t = 0,\ \ p = 0,\ \ \vec{D}_{\cdot,k}(0) = (x,y)$
if $I_{\cdot,k}[x,y] == 1,$
  while $\vec{D}_{\cdot,k}(t) \neq NULL,$
  $(i,j) = \vec{D}_{\cdot,k}(t)$
  $t = t + 1$
  for $m = 0$ *to* $7,$
    if $\left((I_{\cdot,k}[(i,j) + \vec{N}(m)] == 1\right)$
      and $(O[(i,j) + \vec{N}(m)] == 0)),$
    $p = p + 1$
    $\vec{D}_{\cdot,k}(p) = ((i,j) + \vec{N}(m))$
    $O[(i,j) + \vec{N}(m)] = 1,$
    end if
  end for
  end while
end if
$S_{\cdot,k}[x,y] = t.$

$O[x,y]$ is a binary value to determine whether a particular $I_{\cdot,k}[x,y]$ value has been counted previously. $\vec{D}$ is an array of spatial coordinates of *valid coefficients* that support the current coefficient $I_{\cdot,k}[x,y]$. $\vec{N}$ is a set of vectors corresponding to neighboring coefficients.

REFERENCES

[1] C. S. Burrus, R. A. Gopinath, and H. Guo, *Introduction to Wavelets and Wavelet Transforms, A Primer*. Englewood Cliffs, NJ: Prentice-Hall, 1998.

[2] Y. Chou, *Probability and Statistics for Desicision Making*. New York: Holt, Rinehart, and Winston, 1972.

[3] I. Daubechies, *Ten Lectures on Wavelets*. Philadelphia, PA: Soc. Ind. Appl. Math., 1992.

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

[5] ——, "Adapting to unknown smoothness via wavelet shrinkage," *J. Amer. Stat. Assoc.*, vol. 90, pp. 1200–1224, 1995.

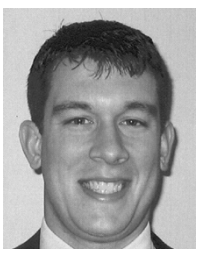
[6] R. Dugad and N. Ahuja, "Video denoising by combining Kalman and Wiener estimates," in *Proc. IEEE Int. Conf. Image Processing*, vol. 4, 1999, pp. 152–156.

[7] F. Faghih and M. Smith, "Combining spatial and scale-space techniques for edge detection to providee a spatially adaptive wavelet-based noise filtering algorithm," *IEEE Trans. Image Process.*, vol. 11, no. 9, pp. 1062–1071, Sep. 2002.

[8] M. Ghazel, G. H. Freeman, and E. Vrscay, "Fractal-wavelet image denoising," in *Proc. IEEE Int. Conf. Image Processing*, vol. 1, 2002, pp. I836–I839.

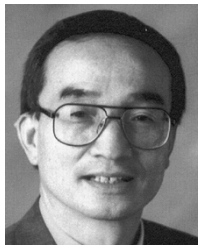
[9] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*. Reading, MA: Addison-Wesley, 1993.

[10] T. C. Hsung, D. Pak-Kong Lun, and W. C. Siu, "Denoising by singularity detection," *IEEE Trans. Signal Process.*, vol. 47, no. 11, pp. 3139–3144, Nov. 1999.
[11] S. J. Huang, "Adaptive noise reduction and image sharpening for digital video compression," in *Proc. IEEE Int. Conf. Computational Cybernetics and Simulation*, vol. 4, 1997, pp. 3142–3147.
[12] C. R. Jung and J. Scharcanski, "Adaptive image denoising in scale-space using the wavelet transform," in *Proc. XIV Brazilian Symp. Computer Graphics and Image Processing*, 2001, pp. 172–178.
[13] S. D. Kim, S. K. Jang, M. J. Kim, and J. B. Ra, "Efficient block-based coding of noise images by combining pre-filtering and DCT," in *Proc. IEEE Int. Symp. Circuits and Systems*, vol. 4, 1999, pp. 37–40.
[14] W. Ling and P. K. S. Tam, "Video denoising using fuzzy-connectedness principles," in *Proc. Int. Symp. Intelligent Multimedia, Video, and Speech Processing*, 2001, pp. 531–534.
[15] W. S. Lu, "Wavelet approaches to still image denoising," in *Proc. Alisomar Conf. Signals, Systems, and Computers*, vol. 2, 1998, pp. 1705–1709.
[16] M. Malfait and D. Roose, "Wavelet-based image denoising using a Markov random field a priori model," *IEEE Trans. Image Process.*, vol. 6, no. 4, pp. 549–565, Apr. 1997.
[17] S. Mallat and W. L. Hwang, "Singularity detection and processing with wavelets," *IEEE Trans. Inf. Theory*, vol. 38, no. 2, pp. 617–623, Mar. 1992.
[18] M. Meguro, A. Taguchi, and N. Hamada, "Data-dependent weighted median filtering with robust motion information for image sequence restoration," *IEICE Trans. Fund.*, vol. 2, pp. 424–428, 2001.
[19] O. Ojo and T. Kwaaitaal-Spassova, "An algorithm for integrated noise reduction and sharpness enhancement," *IEEE Trans. Consum. Electron.*, vol. 46, no. 5, pp. 474–480, May 2000.
[20] R. A. Peters, "A new algorithm for image noise reduction using mathematical morphology," *IEEE Trans. Image Process.*, vol. 4, no. 5, pp. 554–568, May 1995.
[21] A. Pizurica, W. Philips, I. Lemahieu, and M. Acheroy, "A joint inter- and intrascale statistical model for Bayesian wavelet based image denoising," *IEEE Trans. Image Process.*, vol. 11, no. 5, pp. 545–557, May 2002.
[22] P. Rieder and G. Scheffler, "New concepts on denoising and sharpening of video signals," *IEEE Trans. Consum. Electron.*, vol. 47, no. 8, pp. 666–671, Aug. 2001.
[23] A. Said and W. A. Pearlman, "A new, fast, and efficient image codec based on set partitioning in hierarchical trees," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, no. 4, pp. 243–250, Jun. 1996.
[24] L. Shutao, W. Yaonan, Z. Changfan, and M. Jianxu, "Fuzzy filter based on neural network and its applications to image restoration," in *Proc. IEEE Int. Conf. Signal Processing*, vol. 2, 2000, pp. 1133–1138.
[25] H. Stark and J. Woods, *Probability, Random Processes, and Estimation Theory for Engineers*. Englewood Cliffs, NJ: Prentice-Hall, 1994.
[26] A. De Stefano, P. R. White, and W. B. Collis, "An innovative approach for spatial video noise reduction using a wavelet based frequency decomposition," in *Proc. IEEE Int. Conf. Image Processing*, vol. 3, 2000, pp. 281–284.
[27] C. Vertan, C. I. Vertan, and V. Buzuloiu, "Reduced computation genetic algorithm for noise removal," in *Proc. IEEE Conf. Image Processing and Its Applications*, vol. 1, Jul. 1997, pp. 313–316.
[28] Y. F. Wong, E. Viscito, and E. Linzer, "Preprocessing of video signals for MPEG coding by clustering filter," in *Proc. IEEE Int. Conf. Image Processing*, vol. 2, 1995, pp. 2129–2133.
[29] Y. I. Wong, "Nonlinear scale-space filtering and multiresolution system," *IEEE Trans. Image Processing*, vol. 4, no. 6, pp. 774–786, Jun. 1995.
[30] Y. Xu, J. Weaver, D. Healy, and J. Lu, "Wavelet transform domain filter: A spatially selective noise filtration technique," *IEEE Trans. Image Processing*, vol. 3, no. 11, pp. 747–758, Nov. 1994.
[31] V. Zlokolica, W. Philips, and D. Van De Ville, "A new nonlinear filter for video processing," in *Proc. IEEE Benelux Signal Processing Symp.*, vol. 2, 2002, pp. 221–224.



Eric J. Balster (S'02–M'04) received the B.S. and M.S. degrees in electrical engineering from the University of Dayton, Dayton, OH, in 1998 and 2000, respectively, and the Ph.D. degree in electrical engineering from The Ohio State University, Columbus, in 2004.

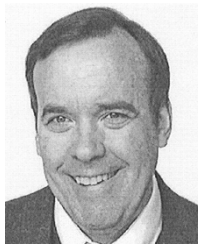
In 2002, he joined the Air Force Research Laboratory's Information Directorate, Wright-Patterson Air Force Base, Dayton. Currently, he is continuing his research in the areas of image and video compression and noise removal in image and video signals.



Yuan F. Zheng (S'82–M'86–SM'90–F'97) received the B.S. degree from Tsinghua University, Beijing, China in 1970, and the M.S. and Ph.D. degrees in electrical engineering from The Ohio State University, Columbus, in 1980 and 1984, respectively.

From 1984 to 1989, he was with the Department of Electrical and Computer Engineering at Clemson University, Clemson, SC. Since August 1989, he has been with The Ohio State University, where he is currently Winbigler Professor and was the Chairman of Electrical and Computer Engineering Department from 1993 to 2004. His research interests include two aspects. One is in wavelet transform for image and video compression for internet and satellite communications. His current efforts focus on content-based compression, 3-D wavelet transformation, video object tracking, and content-based retransmission in Internet communications. The other is in robotics, which includes robots for biological applications, multiple robots coordination, legged robots, human-robot coordination, and personal robotics.

Prof. Zheng was Vice-President for Technical Affairs of the IEEE Robotics and Automation Society from 1996 to 1999. He was an Associate Editor of the IEEE TRANSACTIONS ON ROBOTICS AND AUTOMATION between 1995 and 1997. He was the Program Chair of the 1999 IEEE International Conference on Robotics and Automation, Detroit, MI, May 10–15, 1999. He is currently on the Editorial Board of *International Journal of Multimedia Tools and Applications*, *Autonomous Robots*, the *International Journal of Intelligent Control and Systems*, and the *International Journal of Control, Automation, and Systems*, and he is an Associate Editor of the *International Journal of Intelligent Automation and Soft Computing*. He received the Presidential Young Investigator Award in 1986.



Robert L. Ewing (S'77–M'87–SM'02) received the B.S.E.E. degree and the M.S. degree in physics from the University of Cincinnati, Cincinnati, OH, and the Ph.D. degree in electrical engineering from the University of Dayton, Dayton, OH.

He began his career in the Propulsion Laboratory, Wright-Patterson Air Force Base, Dayton, during the early 1970s with the development of jet engine control systems and the initial control system used on the F-15 aircraft. In the mid-1970s, he was with the University of Cincinnati's Medical School, where he worked in the area of electronic control and regeneration of peripheral (sciatic) nerves used in walking. From 1977 to 1982, he held the position of Medical Research Scientist at the Aerospace Medical Research Laboratory, in the Biodynamic Effects Division. He worked to develop the pilot's analog and digital flight control interfacing and aircraft ejection systems for low-level, high-speed flight. In 1982, he became an Instructor for the Army at the Air Force Institute of Technology (AFIT) and an Adjunct Instructor at Wright State University, Dayton. During his work at the AFIT, he developed many of the early short courses and classes in robotics, digital control, artificial intelligence, neural nets, database systems, low observables (radar), navigation and guidance systems, microprocessor design, and microelectromechanical devices. In 1993, he joined Wright Laboratory's Solid-State Electronic Devices Directorate, working in the area of hardware description language (VHDL) for VLSI synthesis. Currently, he is the Technical Advisor for the Information Directorate's Embedded Information Systems Branch, as well as the Director of the Computer Engineering Research Consortium of local universities in the area of embedded system design (since 1996). He is working toward the development and use of bio-inspired intelligent information processing and data compression for embedded systems and synthesis. He is currently an Adjunct Professor at AFIT and Wright State University.

Dr. Ewing has been a registered Professional Engineer with the State of Ohio since 1984.