Softening the Multiscale Product Method for Adaptive Noise Reduction

Jun Ge and Gagan Mirchandani
Department of Electrical and Computer Engineering
The University of Vermont
Burlington, Vermont 05405, USA

***Abstract*— The goal of denoising is to remove the noise while preserving the important features as much as possible. By exploring the power of parsimonious wavelet basis representation and statistical decision methods, Donoho and Johnstone [5] pioneered the *wavelet shrinkage*. However, the performance of traditional *wavelet shrinkage* is not even as good as that of a simple *multiscale product method* (MPM) [22], because the wavelet basis representation in the traditional *wavelet shrinkage* is not shift-invariant. We numerically reveal the connection between the simple MPM [22] and Donoho-Johnstone's hard thresholding [5]. Based on the observations and an analysis of the MPM, we propose a softened version of MPM which is in analogous to Donoho-Johnstone's soft thresholding [5]. Thanks to the explicit detection of singularities and the use of both ℓ_2 and ℓ_0 stopping criteria to reduce the false detection, the performance of the softened MPM is superior to other methods with redundant wavelet representations for the functions of one-dimensional piecewise linear class. Combined with the local variance analysis discussed elsewhere, we extend the new method to two-dimensional image denoising.**

I. INTRODUCTION

The recovery of a signal from observed noisy data, while still preserving its important features, continues to remain a fundamentally elusive and challenging problem in signal analysis. Noise is traditionally characterized by high frequency components. Accordingly, Fourier-based methods have been employed for their suppression. Concomitantly, this also reduces the sharpness of the significant features since their components often contain high frequencies. By exploring the power of parsimonious wavelet basis representation and using statistical decision methods, Donoho and Johnstone [5] pioneered the concept of *wavelet shrinkage* for denoising. The general procedure for wavelet shrinkage denoising algorithms consists of the following: *Forward wavelet transform, Modification (shrinkage) of detail coefficients, Inverse wavelet transform*. The intuition behind adopting a wavelet basis representation is that important features are assumed to be characterized by large wavelet coefficients across most of the scales, while most of the noise power is considered confined to several fine scales, thus facilitating the separation of *feature-dominated* coefficients $\{w_f\}$ and *noise-dominated* coefficients $\{w_z\}$. Two thresholding methods are generally employed:

$$\begin{aligned}\eta_\lambda^{hard}(w) &= w \cdot 1(|w| > \lambda)\\ \eta_\lambda^{soft}(w) &= sgn(w) \cdot (|w| - \lambda)_+\end{aligned}$$

where $1(\cdot)$ is the indicator function, w the wavelet coefficient value and λ the threshold parameter. Using an orthonormal wavelet basis decomposition, the hard and soft thresholding methods are indeed closed-form solutions to the following two problems: best n-term approximation problem in the mean-squared error (MSE) sense and the l_1-penalized least squares problem respectively. Methods for selecting the threshold parameter therein, have been proposed for the additive Gaussian white noise (GWN) case. These are based on minimax theory [5], minimizing the Stein unbiased risk estimate [6], cross validation [15], and Bayesian approach [1].

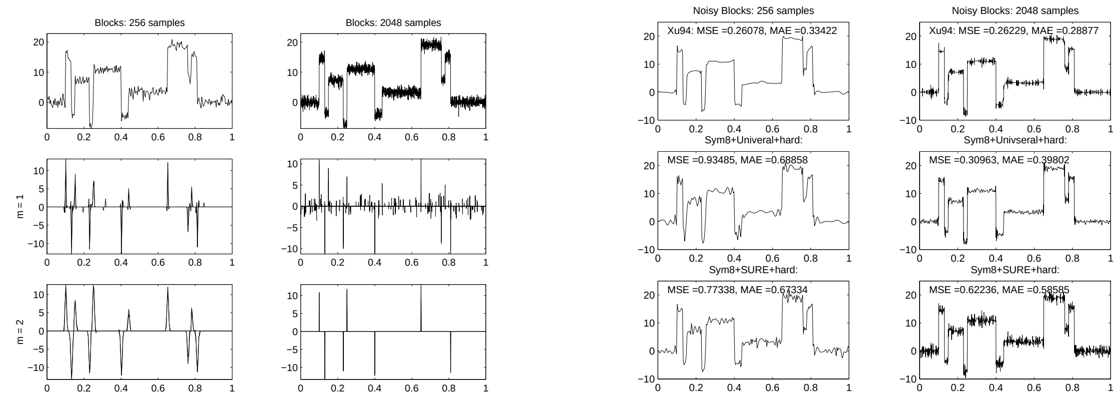
In parallel with Donoho-Johnstone's work, Mallat and Zhong [13] were the first to introduce the complete multiscale edge representation of signals using quadratic spline wavelets. They showed that multiscale edges can be detected and characterized from the local maxima of the wavelet transform. That idea was adopted by Xu *et al.* [22] in their basic and simple algorithm for removing additive GWN from signals. Instead of calculating the Lipschitz exponents [13] to identify edges, Xu *et al.* proposed using correlation of wavelet coefficients across adjacent scales to distinguish significant edges from noise and named their technique *spatially selective noise filtration*. Pan *et al.* [16] examined the behavior of additive GWN in Xu *et al.'s* method and proposed a noise level estimation technique. However, for small size data, the performance of their estimate decreases as the scale becomes coarser. In addition, their coarse-to-fine strategy has difficulty with edges in close proximity. Thus their method avoids extracting noise as edges at fine scales at the price of missing small features. A method similar to Xu *et al.'s* was used by Sadler and Swami [18] and some theoretical justification was provided. They named the basic technique *multiscale product method* (MPM).

The MPM used in our approach is an extension of the spatially selective noise filtration technique proposed by Xu *et. al.* [22]. The reason we adopt the name *multiscale product method* here is that it explicitly points to the use of correlation across scales. We shall discuss the connections between this method and Donoho-Johnstone's hard-thresholding idea. The recently proposed *wavelet footprints* [7] technique attempts to combine the detection and representation of singularities via precise modelling of the discontinuities. Inspired by these observations, we propose in our method a robust noise power estimator as also a soft-thresholding-like technique. The latter avoids the extraction of noise as features while still retaining the weak features. The new method is like a softened version of Xu *et. al.*'s method in that it is similar to soft thresholding.

II. TRADITIONAL MPM AND HARD THRESHOLDING

Using the products of coefficients across scales for image analysis is not a new idea. Indeed, it has appeared in the computer vision society literature even before the advent of wavelets [17]. Witkin provided the foundation for scale space theory [21] by generalizing Rosenfeld's work [17], in which smoothing filters at dyadic scales were used. Based essentially on forming multiscale pointwise products of smoothed gradient estimates, this approach attempts to enhance the peaks of the gradients caused by true edges, while suppressing false peaks due to noise. Thus the wavelet transform must act as an "edge detector", i.e., the detail coefficients should be equivalent to the estimated gradients. In [13] and [22], quadratic spline wavelets are used to play this role. The justification for this transform acting as as a good gradient estimator can be found in [8].

To distinguish edge maxima from noise maxima, Mallat and Zhong [13] analyze the singularity properties of wavelet transform



(a) The selected coefficient from the first two fine scales using the traditional multiscale product method [22]

(b) Reconstructed signals using traditional multiscale product method [22], Universal hard thresholding [5], and SUREShrink hard thresholding [6]

Fig. 1. Traditional multiscale product method and hard thresholding

domain maxima across various scales. Xu *et. al.*'s method relies on the variations in scale of the wavelet transform data, but direct multiplication of wavelet transform data at adjacent scales is used to distinguish important edges from noise. (See [22] for the detailed algorithm). Fig. 1(a) shows the selected detail coefficients using the traditional MPM [22] where we assume that the noise level is known. The signals are the standard *Blocks* signal generated by routines in *Wavelab802*[6] with different number of samples. The signal/noise ratio is taken as 7 in both cases. Fig 1(b) displays the reconstructed signal using the traditional MPM, hard thresholding employing the Universal threshold and hard thresholding using the SURE threshold. We observe the following:

(1) The multiscale product method is equivalent to hard thresholding in that unselected coefficients are removed. The process used by Xu *et al.* for terminating iterations for coefficient selection is equivalent to choosing level-dependent threshold which is the technique employed in SUREShrink [6].

(2) The traditional multiscale product method is better than SUREShrink both quantitatively (MSE) and qualitatively (quality bands and mean absolute error (MAE)). This benefits in part from the fact the undecimated wavelet transform has the shift-invariant property, while the other two use a critically sampled wavelet transform. Other denoising methods with a shift-invariant property, such as translation-invariant denoising by Coifman and Donoho [4] and an undecimated wavelet denoising (UWD) [12] technique were proposed a little later. These effectively extend the Universal and SUREshrink techniques.

(3) Even with a known noise level, Xu *et. al.*'s method cannot avoid extracting noise as features at fine levels, especially for the large size case (2048 samples).

III. SOFTENING THE MULTISCALE PRODUCT METHOD

In this section, we discuss the behavior of noise across scales. Combining this with the observations in the previous section, and also an analysis of the MPM, we generate the structure of our new method.

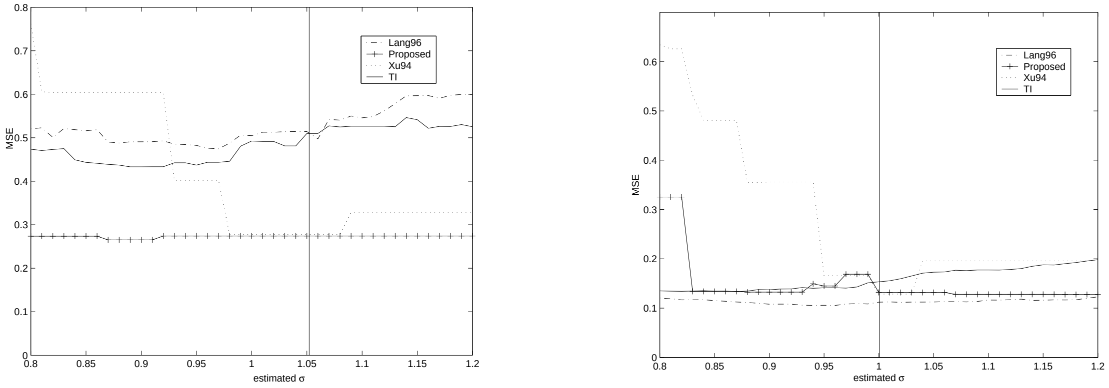
A. Behavior of σ of GWN across Scales

The estimation of noise variance is central to the traditional MPM. Fig. 2 shows the MSE performance dependence of different non-orthogonal wavelet based denoising methods, including the proposed one, with respect to the estimated noise level. Theoretical results for orthogonal wavelet denoising can be found in [2]. In [22], Xu *et al.* proposed using the background noise in the "dark" (signal-free) regions near boundaries for estimation of the reference noise. However, this is not always possible in practice. Even when this is so, other assumptions such as the original signal being constant and availability of a sufficient number of samples need to be made. In [16], the noise level σ is estimated from the first two scales, employing the assumption that the noise dominates at fine scales. Furthermore, it is implicitly assumed that the power of outliers in the noise cannot be larger than about 5 percent of total noise power. Finally, a set of empirical parameters are used.

We now discuss briefly the behavior of noise across scales and examine in particular, the GWN case. In a stochastic setting, a useful characteristic of the orthogonal wavelet transform in contrast to the undecimated wavelet transform, is that the wavelet coefficients of a white noise process is still white. This is simply due to the orthogonality. In their implementation, Donoho and Johnstone obtained an estimate of the noise level from the finest scale coefficients using the median absolute deviation (MAD):

$$\hat{\sigma} = median(|w_{J-1,k|} : 0 \le k < 2^{J-1})/0.6745.$$

The underlying belief here was that it was important to use a robust estimator like the median, in case the fine-scale wavelet coefficients contained a small proportion of strong "signals" mixed in with "noise" [6]. The relative error performance of this estimator of additive WGN is illustrated in Fig. 3. Estimate of $\hat{\sigma}/\sigma_{in}$ for different signals using three different wavelets, are all within 10 percent of the true noise level. Donoho's robust method for estimation of noise level is extensively used in other denoising methods such as those in [4] and [12]. When the undecimated quadratic spline wavelet is used, the highpass analysis filter is same as that for the Haar case. Interestingly, estimation with this wavelet is as good as that with other wavelets with a higher number of vanishing moments. In addition, given additive GWN noise $z \sim \mathcal{N}(0, \sigma^2)$, we know that the variance of the convolution output $z_h = h * z$, where h is an arbitrary FIR filter, can be derived as



(a) 256 samples

(b) 2048 samples

Fig. 2. The dependence of the performance on estimated noise level

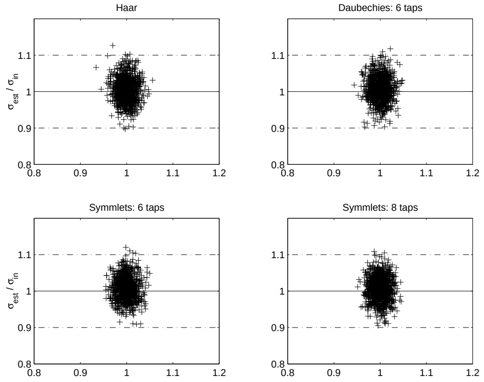


Fig. 3. Performance of MAD for Various Filters

$$Var\{z_h\} = Var\{h * z\} = Var\{z\} \cdot \|h\|_{\ell^2}^2 = \sigma^2 \cdot \|h\|_{\ell^2}^2.$$

Experimental results suggest the use of this formula for no more than three levels depending on the noise level and sample size. Fortunately, the performance of the proposed method is robust to a wide range of estimation errors.

B. Analysis and the New Method

In [18] the multiscale product method is characterized statistically and its performance for detection of position and model parameter estimation of edges is evaluated. It is shown there that multiscale product method generally reduces correlation in the input noise, and that the noise has a heavy tailed distribution. Unfortunately, the determination of the *pdf*s of multiscale products is difficult, even with the additive GWN assumption for the input noisy signal, where a closed form is only available for the bivariate case.

In a deterministic setting, the multiscale product method can be viewed as an extension of Tikhonov regularized differentiation [19] in which the stabilizing operator is chosen as the second order derivative operator d^2/dx^2 and the norm taken as L_2. If we assume that the noiseless signal is bandlimited, the choice of the operator and the norm corresponds to an *a priori* constraint on the function [19]. This leads to the problem of finding some functional $S(x)$ such that

$$\sum_{k=1}^{n} (f_k - S(x_k))^2 + \lambda \int |S''(x)|^2 dx$$

is minimum. The solution of this variational problem has been proven by both Schoenberg and Reinsch as the approximating cubic splines [19], which theoretically justifies the use of the cubic spline function as the scaling function in the MPM. In denoising via the latter, the regularization parameter λ is chosen in a level-dependent fashion. As stated earlier that Donoho-Johnston's hard thresholding and soft thresholding results are the closed-form solution to the optimization problems: minimizing the mean-squared error and the ℓ_1-penalized least squares. In other words, hard thresholding takes the first term of the Tikhonov regularization problem, and soft thresholding changes the second term as ℓ_1 norm of wavelet coefficients. (More detail of this deterministic analysis will be discussed in [9].)

So far, we have discussed the relation between traditional multiscale product method and hard thresholding. We observed that, at fine scales, some of the noise can be falsely extracted as features. Especially for the larger sample signals, the probability of false extraction is high. This can be observed in Fig. 1(a). Based on these observations and the theoretical analysis of multiscale product method, it would be intriguing to find a "softy" sibling for the traditional multiscale product method. This intuition, inspired by soft thresholding, can be explained as follows: Due to the approximation power of (bi-)orthogonal wavelets [20], i. e., the fast decay property of wavelet coefficients, for many noiseless objects such as functions in certain smoothness classes, the wavelet representation is very sparse and contains many zero coefficients. After contamination with noise, the zero wavelet coefficients are not zero anymore; thus the nonzero coefficients are more or less corrupted. Reconstruction using the corrupted coefficients will cause an annoying visual appearance. This explanation is similar to Donoho and Johnstone's arguments for VisuShrink [5]. We propose a *softened* iteration stop-decision for the new multiscale product method. This is as follows:

Softened Stop-decision Criteria:

$$Iterate\ until\ \{(PW(m) < T_{\ell_2})\ \&$$
$$(Max(abs(w(m,n))) < T_{\ell_0})\},$$

where $w(m,n)$ is the detetail coefficient, $PW(m)$ is power of

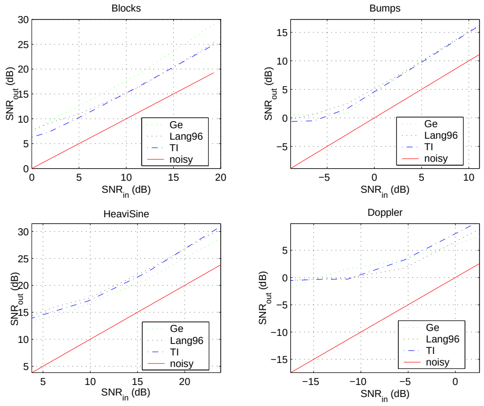


Fig. 4. Comparison for different classes of signals: Blocks, Bumps, HeaviSine and Doppler

the multiscale product of coefficients, T_{ℓ_2} and T_{ℓ_0} are the ℓ_2 and ℓ_0 thresholds respectively. The formulation T_{ℓ_2} is same as that in the traditional multiscale product method (see [22] for more detail). T_{ℓ_0} is same as the universal threshold in [12]. Both the two thresholds both are dependent on the estimation of the noise level. The comparison of the dependence of algorithm performance on noise estimates is shown in Fig. 2. We see that the proposed method has the least dependence on the noise level estimation. In the spirit of Donoho-Johnstone's wavelet thresholding, the proposed method is straightforward, easy to implement, and still theoretical tractable. Fig. 4 displays the performance of various shift-invariant denoising methods using signals from different smoothness classes. For *Blocks* which is comprised of piecewise constant segments, our technique is on average 2.5 dB better. All methods have similar performance for *Bumps*. For signals like *HeaviSine* the new method does not perform as well as other techniques where smoother wavelets with higher number of vanishing moments are used. Note that the quadratic spline has only one vanishing moment. We leave the discussion for the *Doppler* signal to the next section.

IV. DISCUSSION

A. Singularities, Multiscale Product Method, and Footprints

The theme to which this work belongs, is fundamentally the capture and representation of singularities in signals, since most of a signal's information is often carried in the local and irregular structures. Mathematically, these structures manifest themselves as local regularities (or the reverse, singularities) - often measured with Lipschitz exponents. Detection of singularities using undecimated wavelets was first proposed by Mallat and Hwang [14]. They proved that the local Lipschitz exponents of a signal can be estimated by tracing the evolution of its wavelet transform modulus maxima across scales within the so-called *cone-of-influence*. With the assumption that "noiseless" signals have singularities with positive Lipschitz exponents while the noise creates singularities whose Lipschitz exponents are negative, the denoising problem can be effectively solved. However, were the signal to oscillate rapidly in the neighborhood of a singular point, then it is not possible to characterize its Lipschitz regularity from the behavior of the wavelet transform in the cone-of-influence of the singular point. This fact explains the unusual behavior of our method for the *Doppler* signal which has a fast oscillation section.

To avoid the complexity of computing Lipschitz exponents, Xu *et al.* used the multiscale product method. Hsung *et al.* explored the cone-of-influence structure explicitly by computing the sum of the modulus of its wavelet coefficients inside the cone. By using a carefully selected threshold, they select wavelet coefficients for reconstruction that correspond to the regular part of the signal. The recently introduced notion "footprints" by Dragotti and Vetterli [7] attempts to combine both the detection and the representation of singularities. Footprints are traces left by time domain singularities in the wavelet domain. Thus a footprint is a vector containing all the significant wavelet coefficients generated by a singularity. In signal denoising, as a first step, each possible position of discontinuity needs to be given by an oracle or detected by some means. Then using the footprints dictionary $\mathcal{D}$, noisy wavelet coefficients corresponding to each position of discontinuity are projected on the right subspace. If the projection is large enough it is kept, otherwise it is cancelled (thresholding step). In the singularity representation, the idea of footprint is equivalent to the sum of the modulus of its wavelet coefficients inside the cone-of-influence. Indeed, they have very similar denoising algorithms. Alternatively, we may conclude that the footprint method is equivalent to the multiscale product method in that a vector threshold is applied to the set of noisy coefficients rather than a scalar threshold applied to all coefficients in the traditional approach. However, it is clear that the behavior of *footprints* is more theoretically tractable. There are two drawbacks to the foorprints-based denoising algorithm: First, the detection of the location of singularity itself is a difficult problem. Secondly, the assumption that two singularities need to be far enough will in practice generate some restrictions in usage. The multiscale product method is more flexible with respect to these two concerns.

B. Extension to 2D

Unfortunately, like the MPM, footprints-based denoising does not work well for 2D images. Dragotti and Vetterli proposed the 2D extension – edgelets [7]. This method does not seem very satisfying in practice. Recently, a simple and efficient image denoising algorithm based on the multiscale product method combined with local covariance analysis has been proposed [8]. The algorithm adaptively weighs the joint inter- and intrascale statistics of detail coefficients. Direct correlation of detail coefficients across scales is used to select the significant coefficients. Intrascale statistics are used to adaptively modify the coefficients, using a new geometric measure. Unlike existing algorithm using parametric models, prior knowledge and estimation of parameters are not needed. Implementation is simple and efficient, with a performance comparable to results by state-of-art methods. Comparison with other denoising techniques is illustrated in Fig. 5. (See [8] for more detail.)

V. CONCLUSIONS

In this paper, a new method was introduced for adaptive noise reduction. The connection between the simple MPM [22] and Donoho-Johnstone's hard thresholding [5] was illustrated numerically. Based on the observations and an analysis of the MPM, we proposed a softened version of the latter, which is in analogous to Donoho-Johnstone's soft thresholding [5]. The detection of singularities in denoising problem is critical because they contain important information. We discussed several different treatments to singularities. Our new method is simple because the detection does not rely on the calculation of Lipschitz exponents which is computationally expensive. Furthermore, the complexity is exactly the same as that of the traditional MPM [22]. It is also preferable because of the better way to reduce false detection of important features. Our current investigation is focused on the combination of local phase

Fig. 5. Comparison for 2D images

information, *e. g., phase congruency* [11], with the current used local magnitude information in the detection and representation of singularities for signal and image processing tasks.

ACKNOWLEDGMENT

Research by both authors is supported in part by DEPSCoR Grant No. F49620-00-1-0280.

REFERENCES

[1] F. Abramovich, T. Sapatinas, and B. W. Silverman, “Wavelet thresholding via a Bayesian approach,” *J. R. Statist. Soc. B*, vol. 60, no. 4, pp. 725-749, 1998.

[2] A. G. Bruce, and H.-Y. Gao, “Understanding WaveShrink: Variance and bias estimation,” *Biometrika*, vol. 83, no. 4, pp. 727-745, Dec. 1996.

[3] A. Chambolle, R. A. DeVore, N.-y. Lee, and B. J. Lucier, “Nonlinear wavelet image processing: Variational problems, compression, and noise removal through wavelet shrinkage,” *IEEE Trans. Image Processing*, vol. 7, no. 3, pp. 319-335, Mar. 1998.

[4] R. Coifman, and D. Donoho, “Translation Invariant De-noising”, *Lecture Notes in Statistics 103*, New York: Springer, pp. 125-150, 1995.

[5] D. L. Donoho and I. M. Johnstone, “Ideal spatial adaptation by wavelet shrinkage,” *Biometrika*, vol. 81, pp. 425-455, 1994.

[6] D. L. Donoho and I. M. Johnstone, “Adapting to unknown smoothness via wavelet shrinkage,” *J. Amer. Statist. Assoc.*, vol. 90, pp. 1200-1224, Dec. 1995.

[7] P. L. Dragotti, and M. Vetterli, “Wavelet footprints: Theory, algorithms, and applications ,” *IEEE Trans. Signal Processing*, vol. 51, no. 5, pp. 1306-1323, May 2003.

[8] J. Ge, and G. Mirchandani, “A simple and efficient wavelet-based denoising algorithm using joint inter- and intrascale statistics adaptively,” *ISSPA*, Paris, France, 2003.

[9] J. Ge, and G. Mirchandani, “A restricted nonlinear approximation method for image denoising based on multiscale product method and local covariance analysis,” in preparation.

[10] T.-C. Hsung, D. P.-K. Lun, and W.-C. Siu, “Denoising by singularity detection,”*IEEE Trans. Signal Processing*, vol. 47, no. 11, pp. 3139-3144, Nov. 1999.

[11] P. D. Kovesi, “Invariant measure of image features from phase information,” PhD Thesis, The University of Western Australia, May 1996.

[12] M. Lang, H. Guo, J. E. Odegard, C. S. Burrus, and R. O. Wells, Jr., “Noise reduction using an undecimated discrete wavelet transform,” *IEEE Signal Processing Letters*, vol. 3. no. 1, pp. 10-12, Jan. 1996.

[13] S. G. Mallat and S. Zhong, “Characterization of signals from multiscale edges,” *IEEE Trans. PAMI*, vol. 14, no. 7, pp. 710-732, July 1992.

[14] S. G. Mallat, and W. L. Hwang, “Singularity detection and processing with wavelets,” *IEEE Trans. Inform. Theory*, vol. 38, no. 2, pp. 617-643, Mar. 1992.

[15] G. P. Nason, “Wavelet shrinkage using cross validation,” *Journal of the Royal Statistical Society, Series B*, vol. 58, no. 2, pp. 463-479, 1996.

[16] Q. Pan, L. Zhang, G. Dai, and H. Zhang, “Two denoising methods by wavelet transform,” *IEEE Trans. Signal Processing*, vol. 47, no. 12, pp. 3401-3406, Dec. 1999.

[17] A. Rosenfeld, “A nonlinear edge detection technique,” *Proc. IEEE*, vol. 58, pp. 814-816, May 1970.

[18] B. M. Sadler and A. Swami, “Analysis of multiscale products for step detection and estimation,” *IEEE Trans. Inform. Theory*, vol. 45, no. 3, pp. 1043-1051, Apr. 1999.

[19] V. Torre and T. A. Poggio, “On edge detection,” *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-8, no. 2, pp. 147-163, Mar. 1986.

[20] M. Unser, “Approximation power of biorthogonal wavelet expansions,” *IEEE Trans. Signal Processing*, vol. 44, no. 3, pp. 519- 527, Mar. 1996.

[21] A. Witkin, “Scale-space filtering,” *Proc. Int. Joint Conf. Artif. Intell.*, pp. 1019-1021, 1983.

[22] Y. Xu, J. B. Weaver, D. M. Healy, Jr., and J. Lu, “Wavelet transform domain filters: a spatially selective noise filtration technique,” *IEEE Trans. Image Processing*, vol. 3, no. 6, pp. 747-758, Nov. 1994.