Wavelet Domain Shrinkage Methods for Noise Removal in Images: A Compendium

E.Jebamalar Leavline
Anna University of Technology, Tiruchirappalli
India

S.Sutha
Anna University of Technology, Tiruchirappalli
Panruti
India

D.Asir Antony Gnana Singh
M.I.E.T Engineering College
Tiruchirappalli
India

ABSTRACT
This paper presents an overview of various threshold methods for image denoising. Wavelet transform based denoising techniques are of greater interest because of their performance over Fourier and other spatial domain techniques. Selection of optimal threshold is crucial since threshold value governs the performance of denoising algorithms. Hence it is required to tune the threshold parameter for better PSNR values. In this paper, we present various wavelet based shrinkage methods for optimal threshold selection for noise removal.

General Terms
Image denoising, Wavelet based methods.

Keywords
Denoising, Spatial domain methods, Wavelet shrinkage, optimal threshold selection

1. INTRODUCTION
In general, an image may be contaminated by noise during acquiring and transmission. The noise present in the images may appear as additive or multiplicative components which have been modelled in a number of ways in the literature [1],[17] such as Gaussian noise, Speckle noise, Salt & Pepper noise, Impulse noise etc... As the occurrence of noisy pixels in the image is random in nature, their distributions are modelled using probabilistic methods [20] [24]. In most of the real time applications such as medical imaging, satellite image data analysis, remote applications etc.., the noisy components have to be removed to ensure faithful information retrieval from the images. A common defect in the imaging system is unwanted non linearity in the sensor and display system. Post processing correction of sensor signals and pre-processing correction of display signals can reduce degradations substantially [1]. Hence pre-processing is essential in any information analysis and retrieval system. Denoising is one of the pre-processing techniques which have drawn much attention of the researchers over a few decades. In this paper, present a detailed survey of various noise removal techniques, with a focus on threshold computing methods is presented since choosing the threshold is crucial in the process of denoising. This paper is organized as follows. Section 2 presents denoising procedure and classification of denoising methods. Section 3 discusses about the wavelet based denoising techniques. Various threshold methods and the tradeoffs involved in selecting an optimal threshold are presented in Section 4. Finally, discussions on observations and conclusion are presented in Section 5.

2. METHODS OF DENOISING
If f(x,y) be the uncorrupted image of size NXN and n(x,y) be the noise function, then the noisy image observation g(x,y) with additive noise is given by

$$g(x,y) = f(x,y) + n(x,y) \qquad \forall\, x,y \le N$$

The process of denoising is nothing but the estimation of the information from noisy observation and may be described as

$$\hat{f}(x,y) = g(x,y) - \hat{n}(x,y)$$

With this background, the state of art denoising methods can be categorized as follows.

2.1 Spatial Filtering Techniques
Spatial filtering is the method of choice in situations when only additive noise is present. This category consists of mean filter and the order statistics filter such as Median filter, Maximum and Minimum filter, Midpoint and Alpha trimmed median filter. Arithmetic and Geometric mean filters are well suited for random noise like Gaussian or uniform noise. The Contra-harmonic filter is well suited for impulse noise, but it requires the prior knowledge about the noise (light or dark). As found in the literature [1],[17], median filter can perform well in removing impulse noise while the number of passes of the median filter has to be kept as low as possible, since larger number of passes may result in blurred images. The process of spatial filtering consists of moving the filter mask (Fig: 1) from point to point in an image. At each point (x,y) the response of filter at that point is calculated. The mask may be of any size of interest (3X3, 5X5, 7X7 etc...). Also, it has to be noted that size of the filter mask affects the performance of the filter [15].

$\frac{1}{9}$ X

1	1	1
1	1	1
1	1	1

$\frac{1}{16}$ X

1	2	1
2	4	2
1	2	1

Figure 1 Spatial Filtering Mask

Another class of filters which fall under spatial filters is adaptive filter, who's behaviour changes based on the statistical characteristics of the image inside the filter region defined by m x n rectangular window. These filters can offer superior denoising performance with the cost of increased complexity [17] [24].

Adaptive median filter is the prime variant of adaptive filter. Filter mask size is altered according to the parameters calculated in the mask area considered originally. It performs well for the impulse noise with low spatial density and seeks to preserve details while smoothing non-impulse noise too. Researchers have shown interest to evolve adaptive iterative median filter which outperforms even for high density noises [26].

2.2 Frequency domain filtering

Frequency domain filtering can be used for periodic noise reduction and removal. This category of filters include band pass filter, band stop filter, Notch (Reject/Pass) filters. The appropriate filter can be chosen with the prior knowledge of noise distribution. The various Fourier domain filtering techniques such as Inverse filter, Wiener filter and least square filter are found in literature. A simple method of removing multiplicative noise like speckle noise too has been proposed namely homomorphic filtering [1] [17]. Fourier transform has been found to be an important image processing tool for image processing and analysis. The major advantage of Fourier domain analysis is that, it can explore the geometric characteristics of a spatial domain image [2]. It has been used for the removal of additive noises from the images. Unlike Fourier transform, Wavelet transform shows localization in both time and frequency and hence it has proved itself to be an efficient tool for a number of image processing applications including noise removal [19]. Fourier transform based methods are less useful because, they cannot work on non-stationary signals and they can capture only global features. But in the real scenario, as the images are only piecewise smooth and the noise distributions are random in nature, Fourier transform cannot perform well for the stochastic noise, but wavelets can do. Hence wavelet based noise removal has attracted much attention of the researchers for several years [4], [6]. A detailed study on wavelet based denoising techniques is presented in the next section

3. WAVELET DENOISING

Wavelet transform is the mathematical tool used for various image processing applications such as noise removal, feature extraction, compression and image analysis. The general method of wavelet based denoising is that, the noisy image may first be transformed to wavelet domain [2] [6].

The transformed image appears as four subbands (A, V, H, and D) as shown in Fig 1 based on the level of decomposition 'j'. 2D discrete wavelet transform leads to decomposition of approximate coefficients at level 'j' into four components i.e. the approximation at level 'j+1' and details in three orientations (Horizontally, Vertically and Diagonally) [25]. Since the noisy components are of high frequency, the three higher bands may contain the noisy components [25], and proper threshold may be applied to smooth the noisy wavelet coefficients followed by the inverse 2D-DWT may be applied to reconstruct the denoised image. Selection of optimal threshold is crucial for the performance of denoising algorithm. Threshold is selected based on the image and noise priors such as mean and variance [10] [23]. Selection of optimal threshold along with various types of wavelet threshold methods is presented in the next section.

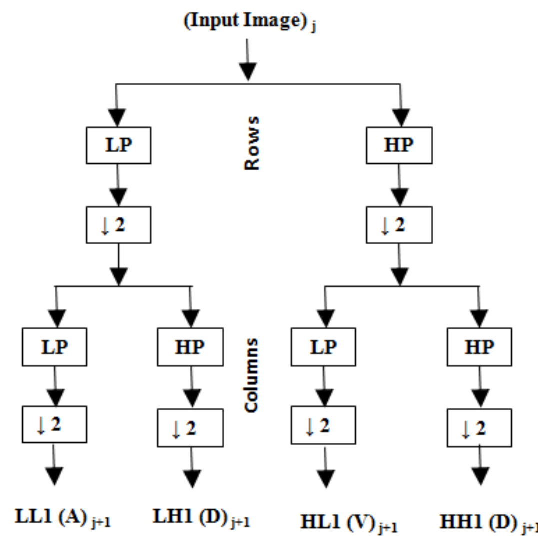

Figure 2 2D Discrete Wavelet Transform First Level Decomposition

(a)

(b)

Figure 3 (a) 2D DWT decomposition (b) 2D DWT decomposition of noisy image (AWGN, zero mean and 0.01 variance)

4. WAVELET BASED THRESHOLD METHODS

4.1 Visushrink

Donoho & Johnston derived Universal threshold and showed that the expected maximum is

$$E_{\max}[N(x)] = \sqrt{2\log_e n}$$

with probability 1 for 'n' independent, identically distributed, standard normal variables with $N(0,\sigma^2)$ .This led to the universal threshold:

$$T_u = \sqrt{2log_e N}\ \hat{\sigma}$$

where 'σ' is an estimate of the population's standard deviation and 'N' is the number of pixels in the image. In practice, 'σ' is calculated as the mean of the absolute difference (MAD) which is more robust than the standard deviation of the sample. The universal threshold method assumes that all wavelet coefficients less than T_u are noise, and these are eliminated. An inherent assumption in this method is that the noise is Gaussian distributed [5] [14].

However, for denoising images, Visushrink is found to yield an overly smoothed estimate. This is because the universal threshold is derived under the constraint that with high probability the estimate should be at least as smooth as the signal. So the universal threshold tends to be high for large values of N, killing many signal coefficients along with the noise. Thus, this threshold does not adapt well to discontinuities in the signal.

4.2 Bayes Shrink

Bayes Shrink has attracted much attention since it sets different thresholds for every subband. Here subbands are frequency bands that differ from each other in level and direction. The relationship between the wavelet transform of the degraded image, uncorrupted image and generalised Gaussian noise with distribution $N(0,\sigma^2)$ (Y, X and V respectively), can be modeled as Y = X+V. Since all the above three factors are mutually independent, their variances are modeled as

$${\sigma_y}^2 = {\sigma_x}^2 + \sigma^2$$

Since huge information about the noise is available at the diagonal coefficients of first level wavelet decomposition (HH1) the noise variance 'σ' is calculated using the robust estimator

$$\hat{\sigma}^2 = \left[\frac{median\ (|X_{ij}|)}{0.675}\right]^2 \quad X_{ij} \in HH1$$

Variance of the corrupted image is

$$\hat{\sigma}_y^2 = \frac{1}{M}\sum_{m=1}^{M} W_m^2$$

Where W_m are the wavelet coefficients in each scale and M is the total number of wavelet coefficients. With this background, the threshold using Bayesshrink is calculated as

$$T_{BS} = \frac{\hat{\sigma}_v^2}{\hat{\sigma}_x} \quad \text{where}$$

$$\hat{\sigma}_x = \sqrt{\max(\hat{\sigma}_y^2 - \hat{\sigma}_v^2,\quad 0)}$$

With $T_{BS} = \max\{|Wm|\}$ for $\hat{\sigma}_v^2 \geq \hat{\sigma}_y^2$ the threshold function may be generalised as follows

$$T_{BS} = \begin{cases} \dfrac{\hat{\sigma}_v^2}{\hat{\sigma}_x} & if\ \hat{\sigma}_v^2 < \hat{\sigma}_y^2 \\ \max\{|Wm|\} & otherwise \end{cases}$$

The Bayesshrink method is effective for images corrupted by Gaussian noise. Bayes shrink is less sensitive to the presence of noise in the areas around the edges [9] [11]. However, the presence of noise in flat regions of the image is perceptually more noticeable by the human visual system. Bayes shrink performs little denoising in high activity sub-regions to preserve the sharpness of edges but completely denoised the flat sub-parts of the image.

4.3 Normalshrink:

This is an adaptive threshold estimation method for image denoising in the wavelet domain based on the generalized Gaussian distribution (GGD) modeling of subband coefficients... The threshold is computed by

$$T_N = \beta\frac{\hat{\sigma}^2}{\sigma_y}$$

where σ and σ_y are the standard deviation of the noise and the subband data of noisy image respectively. β is the scale parameter, which depends upon the subband size and number of decompositions**,** computed as

$$`\beta = \sqrt{log\left(\frac{L_k}{J}\right)}$$

$$\hat{\sigma}^2 = \left[\frac{median\ (|X_{ij}|)}{0.675}\right]^2 \quad X_{ij} \in HH1$$

L_k is the length of the subband at k^{th} level, J is the total number of decompositions, σ^2 is the estimated noise variance of HH1 subband and σ_y is the standard deviation of the image subband. This method is computationally more efficient and adaptive because the parameters required for estimating the threshold depend on subband data. Performance of normal shrink is similar to bayes shrink. But normal shrink preserved edges better than noise removal method using the bayes shrink method as well as removing noise better than bayes shrink [12].

4.4 Modified Bayes shrink

As found in literature [9] [11], noisy components are not sufficiently removed using Bayes shrink. Modified bayes shrink remove noise better than bayes shrink. Thresholds are different for each subband coefficient as shown below.

$$T_{MBS} = \beta\frac{\hat{\sigma}_v^2}{\hat{\sigma}_x}$$

Where

$$\beta = \sqrt{\frac{\log M}{2X\ j}}$$

'M' is the total of coefficients of wavelet. 'j' is the wavelet decomposition level present in the subband coefficients under consideration. The modified bayes shrink yields the best results for denoising and preserves edges better than bayes shrink and normal shrink**.**

4.5 Bivariate shrink:

Bivariate model is used to characterize the dependency between a wavelet coefficient and its parent. This is well suited for wavelet variants such as quaternion wavelet which has strongly Gaussian characteristics. The corresponding bivariate maximum a posterior (MAP) estimator is based on noisy wavelet coefficients [13]. The basic Bayesian estimation method has been modified via considering the dependency between wavelet coefficients and its parent. Let W_2 represent the parent of W_1. Then,

$$y_1 = W_1 + n_1$$
$$y_2 = W_2 + n_2$$

Where n_1 and n_2 are the noisy components, y_1 and y_2 are noisy observations of W_1 and W_2

Hence y = W+ n, where y = (y_1 ,y_2), W = (W_1 ,W_2), n= (n_1, n_2)

The wavelet coefficients (W) may be estimated with the prior knowledge of noisy observation y as

$$\widehat{W}(y) = arg \max_{W} P_{W|y}(w|y) = arg \max_{W}[P_n(y = W)P_W(W)$$

$P_W(W)$ is the probability distribution function of Wavelet coefficients which follows Gaussian pattern and the wavelet coefficients are estimated as,

$$\widehat{W}_1 = \frac{\left(\sqrt{y_1^2 + y_2^2} - \frac{\sqrt{3}\,\sigma_n^2}{\sigma}\right)}{\sqrt{y_1^2 + y_2^2}} + y_1$$

With threshold value TBV = $\frac{\sqrt{3}\,\sigma_n^2}{\sigma}$ the model of bivariate shrinkage function may be defined as BFM = (y_c, y_p, σ_n, σ) , y_c, y_p are the subband and parent wavelet coefficients respectively. σ_n and σ are the noise and marginal variance respectively and they are calculated as in the case of Bayes shrink [22].

4.6 Sure shrink:

Sure Shrink is more explicitly adaptive to unknown smoothness and has better large-sample MSE properties. This method is a subband adaptive threshold scheme, based on Stein's unbiased estimator for risk (SURE) (quadratic loss function) [6-8]. One gets an estimate of the risk for a particular threshold value t. minimizing the risks in 't' gives a selection of the threshold value. Let the transformed coefficients in the 'j'th subband be $\{X_i : i = 1,\ldots,d\}$. SURE proposes method for estimating loss $\left\|\hat{X} - X\right\|^2$. For the soft threshold estimator $\hat{X}_i = \eta_t(X_i)$,$\hat{X}_i = \eta_t(X_i)$, $SURE(t;X) = d - 2\#\{i: |X_i| \leq t\} + \sum_{i=1}^{d} \min(|X_i|)^2$ Then select threshold t^S by

$$t^s = argmin\, SURE(t;X)$$

The SURE principle can be used to select a threshold that is applied to the image data, resulting in an estimate of the mean vector. This estimate is sparse and much less noisy than the raw image data [14]. The SURE principle just described has a serious draw-back in situations of extreme sparsity of the wavelet coefficients. In such cases the noise contributed o the SURE profile by the many coordinates at which the signal is zero, swamps the information contributed to the SURE profile by the few coordinates where the signal is nonzero. Consequently, Sure Shrink uses a Hybrid scheme [16].

4.7 Minimax Threshold

It uses a fixed threshold chosen to yield minimax performance for mean square error against an ideal procedure. The minimax principle is used in statistics in order to design estimators. Since the de-noised signal can be assimilated to the estimator of the unknown regression function, the minimax estimator is the one that realizes the minimum of the maximum mean square error obtained for the worst function in a given set. Minimax threshold does not give good visual quality, but it has the advantage of giving predictive performance [21].

4.8 Waveshrink:

It is an expansion based estimator proposed by Donoho and Johnston. With the orthonormallitty property of the wavelets is has been prove din the literature that the least square estimate of the wavelet coefficients is unbiased and the risk function values are equal to the risk in coefficient values. A mere least square estimate does not denoise the original image [21]. Hence, to estimate wavelet coefficients at minor risk Donoho and Johnston have applied a component wise function which shrinks the least square estimate towards zero. Total risk factor is calculated by summing component wise risks. With this background, the threshold function is modified [21] as

$$T_{WS} = \sqrt{2\log(N\log N)}$$

Where N is the total number of pixels

4.9 Cycle spinning

Continuous and discrete Wavelet transforms lack in shift-invariance property. The goal of cycle-spin thresholding is to include some new significant coefficients in shifted-image transforms, thus producing a sharper denoised image which includes more edge details. Cycle-spin thresholding achieves shift-invariance by averaging all shifts of the noisy image [18]. This will eliminate over smoothing of edges. Every cycle shift is denoised and all these denoising are averaged by simple arithmetic mean. This yields higher SNR than the methods of denoising with shrinkage functions.

4.10 Trade off between Threshold, PSNR and Complexity

Selection of optimal threshold determines the efficiency of the denoising algorithm [10]. The common measure of quality in images in peak signal to noise ratio which may be defined as

$$PSNR = 10\log_{10}\left(\frac{255^2}{MSE}\right)$$

Here MSE is the mean square error whose magnitude quantifies the presence of noise and the performance of denoising algorithm. As discussed in section - IV wavelet based shrinkage algorithms give better estimate of the noise priors and hence the threshold with the expense of high computational complexity. It is very crucial to select the threshold value with less computational complexity and with significant improvements in PSNR.

5. DISCUSSIONS AND CONCLUSION

As discussed in sections 3 and 4, Wavelet transforms have proved themselves to be efficient tool for image denoising. Subband adaptive wavelet shrinkage methods exhibit near optimal estimate of the threshold. Sureshrink and Bayesshrink are found as the basic denoise variants. Normalshrink outperforms Bayesshrink and sureshrink in terms of noise removal performance. Also, the denoising performance depends on the thresholding methods whether it is hard threshold or soft threshold [3]. Selecting optimal threshold for denoising is still an area of thrust for the research community and there is scope for further enhancement of the threshold methods discussed.

6. REFERNECES

[1] Anil K.Jain 2003, "Fundamentals of Digital Image Processing" PHI

[2] Boggess & Narcowich, 2002, "A First Course in Wavelets with Fourier Analysis", Prentice Hall

[3] David L. Donoho 1993 "De-Noising by Soft Thresholding", *IEEE Trans. Info. Theory 43*, pp. 933-936.

[4] David L. Donoho, Iain M. Johnstone, Gérard Kerkyacharian, Dominique Picard 1993 "Wavelet Shrinkage: Asymptopia".

[5] D. L. Donoho and I. M. Johnstone 1994, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, no. 3, pp. 425–455.

[6] David L. Donoho and Iain M. Johnstone 1995,"Adapting to Unknown Smoothness via Wavelet Shrinkage", Journal of the American Statistical Association Vol. 90, No. 432 pp. 1200-1224.

[7] Fengxia Yan, Lizhi Cheng, and Silong Peng 2008, "A New Interscale and Intrascale Orthonormal Wavelet Thresholding for SURE-Based Image Denoising", IEEE Signal Processing Letters, Vol. 15.

[8] Florian Luisier, Thierry Blu, *and* Michael Unser (June 2010), "SURE-LET for Orthonormal Wavelet-Domain Video Denoising", IEEE Transactions On Circuits And Systems For Video Technology, Vol. 20, No. 6.

[9] S.Grace Chang, Bin Yu, Martin Vetterli 2000, "Adaptive wavelet thresholding for denoising and compression", IEEE Trans. On Image processing, Vol.9, No.9, PP. 1532-1546.

[10] Hamed Pirsiavash, Shohreh Kasaei, and Farrokh Marvasti, 2005, "An Efficient Parameter Selection Criterion for Image Denoising", IEEE International Symposium on Signal Processing and Information Technology.

[11] Iman Elyasi, and Sadegh Zarmehi 2009," Elimination Noise by Adaptive Wavelet Threshold" World Academy of Science, Engineering and Technology.

[12] Lakhwinder Kaur, Savita Gupta and R.C. Chauhan, "Image Denoising using Wavelet Thresholding".

[13] Levent Sendur, Ivan W. Selesnick *2002*,"Bivariate Shrinkage Functions for Wavelet-Based Denoising Exploiting Interscale Dependency" IEEE Transactions On Signal Processing, Vol. 50, No. 11.

[14] Leavline, E.J.; Sutha, S 2011, "Gaussian noise removal in gray scale images using fast Multiscale Directional Filter Banks", IEEE International Conference on Recent Trends in Information Technology (ICRTIT-2011) pp 884 - 889

[15] Loupas, T.; McDicken, W.N.; Allan, P.L.(Jan 1989), "An adaptive weighted median filter for speckle suppression in medical ultrasonic images", IEEE Transactions on Circuits and SystemsVolume: 36 Issue:1 pp 129 – 135

[16] Martin Raphan, and Eero P. Simoncelli (August 2008), "Optimal Denoising in Redundant Representations" IEEE Transactions On Image Processing, Vol. 17, No. 8.

[17] Rafael C.Gonzalez, Richard E.Woods 2002, "Digital Image Processing", Second Edition, Pearson Education Asia.

[18] Ramin Eslami and Hayder Radha 2003,"The Contourlet Transform for Image De-noising Using Cycle Spinning", IEEE Trans. Image Processing pp. *1982*-1986.

[19] Rao R M and A S Bopardikar 2000, "Wavelet Transforms Introduction to theory and Applications", Pearson Education, Asia.

[20] Saeed V. Vaseghi 2000, "Advanced Digital Signal Processing and Noise Reduction,", John Wiley & Sons Ltd.

[21] Sardy, S 2000, "Minimax Threshold for Denoising Complex Signals with Waveshrink", IEEE Transactions On Signal Processing, , VOL 48; PART 4, pages 1023-1028

[22] Shan GAI, Peng LIU, Jiafeng LIU, Xianglong TANG, (2010) "A New Image Denoising Algorithm via Bivariate Shrinkage Based on Quaternion Wavelet Transform" , Journal of Computational Information Systems 6:11 3751-3760

[23] S.Sudha, G.R.Suresh, R.Sukanesh 2007, "Wavelet based Image denoising using adaptive thresholding", International conference on Computational Intelligence and Multimedia Applications, PP. 296-300.

[24] Tony F. Chan, Jianhong[Jackie] shen 2005 , "Image Processing and Analysis,", Society for Industrial and Applied Mathematics, Philadelphia.

[25] Vetterli M Kovacevic J 1995, "Wavelets and Sub band Coding", Prentice Hall.

[26] Wei Li; Yanxia Sun; Shengjian Chen 2009, "A New Algorithm for Removal of High-Density Salt and Pepper Noises", IEEE 2nd International Congress on Image and Signal Processing.pp.1-4.