A Comparative Study of Wavelet Thresholding for Image Denoising

Arun Dixit, Poonam Sharma
Department of Computer Science & Engineering, M.I.T.S., Gwalior, India
Email: adixit231@gmail.com, enggpoonam@yahoo.com

Abstract—Image denoising using wavelet transform has been successful as wavelet transform generates a large number of small coefficients and a small number of large coefficients. Basic denoising algorithm that using the wavelet transform consists of three steps—first computing the wavelet transform of the noisy image, thresholding is performed on the detail coefficients in order to remove noise and finally inverse wavelet transform of the modified coefficients is taken. This paper reviews the state of art methods of image denoising using wavelet thresholding. An Experimental analysis of wavelet based methods Visu Shrink, Sure Shrink, Bayes Shrink, Prob Shrink, Block Shrink and Neigh Shrink Sure is performed. These wavelet based methods are also compared with spatial domain methods like median filter and wiener filter. Results are evaluated on the basis of Peak Signal to Noise Ratio and visual quality of images. In the experiment, wavelet based methods perform better than spatial domain methods. In wavelet domain, recent methods like prob shrink, block shrink and neigh shrink sure performed better as compared to other wavelet based methods.

Index Terms—Image Denoising, Wavelet Transform, Wavelet Shrinkage, Wavelet Thresholding, Gaussian Noise Reduction.

1. INTRODUCTION

Image denoising is the method of removing noise acquired at image capturing and transmission state. Initial methods proposed for image denoising were based on statistical filter [1, 18], but the problems associated with spatial filter are that during denoising process low pass filter makes the edges blur and high pass filter amplify noisy background. FFT based denoising method is introduced which is a low pass filtering technique and it is unable to preserve sharpness of the edges as the basis function of FFT is not being localized in terms of time or space domain [6]. This problem can be resolved by wavelet transform as of its localized nature in time & space domain and performs denoising with edge preservation. Mallat [15] has given multi resolution theory of wavelets. Wavelets have various advantages like no redundancy and efficient implementation. There is no redundant information stored, as wavelet functions are orthogonal. Wavelets can be easily implemented using simple algorithms based on convolution. The initial work on wavelet based denoising using thresholding was done by Donoho and Johnstone [2]. Donoho and Johnstone proposed visu shrink which is universal threshold based first attempt for wavelet based denoising. Donoho and Johnstone also proposed sure shrink which uses sub band level thresholding. Later on various wavelet based methods are proposed. Our work aims to study those methods and identify their features.

Wavelet denoising using thresholding involves three basic steps—first step involves computation the wavelet transform of noisy image, second step is used to apply thresholding on noisy wavelet coefficient according to some rule and finally computing inverse wavelet transform of modified wavelet coefficients. Wavelet denoising using thresholding algorithm is also known as wavelet shrinkage in which wavelet coefficient of noisy image are grouped based on certain threshold value and threshold function.

There are three type of threshold value selection method universal threshold, sub band adaptive threshold method, spatially adaptive method. In image denoising, initially universal thresholding is proposed which utilizes same threshold value for all sub bands of image [3]. As all sub bands don’t contain same level of noise so sub band adaptive threshold is proposed [2]. Sub band adaptive threshold has shown better denoising performance as compared to universal threshold. In order to exploit spatial characteristics of image, spatially adaptive threshold is proposed [4]. The basic thresholding functions are hard and soft thresholding as proposed in [2]. Small coefficients are dominated by noise, while coefficients have a large coefficients contain more signal information as compared to noise so both thresholding functions set the coefficients less than threshold to zero. A hard thresholding function doesn’t change rest of coefficients while soft thresholding reduces its value by absolute threshold value.

In last two decades, various wavelet based methods are proposed like Universal threshold methods, sub band adaptive threshold methods, block level threshold methods and spatially adaptive threshold methods etc. Mathematical model are also introduced like Bayesian modeling. Another important aspect of wavelet based denoising is estimation of variance. A good variance estimator is median of absolute deviation (MAD) which is computed using diagonal coefficients of first
decomposition of image. In this paper, various properties of wavelet domain methods are analyzed so that essential characteristics of a good denoising algorithm could be obtained. These all state of art algorithm introduces some novel approaches. This work investigates those approaches like way of thresholding, estimation of variance and exploiting spatially adaptive characteristics.

The remainder of the paper is structured as follows. Section II deals with wavelet transform, its decomposition process and its families. Section III describes wavelet based denoising process in which detail coefficients are thresholded. Various thresholding functions and thresholding values are also explained in this section. Section IV introduces various wavelet domain image shrinkage methods which could be considered as state of art methods for image denoising. In section V, experiment is performed on MATLAB simulation environment and an analysis of these state of art methods is done on the basis of visual quality and Peak Signal to Noise Ratio (PSNR) and section VI concludes the paper.

II. WAVELET TRANSFORM

The wavelet Transform provides good time and poor frequency resolution at high frequencies, while at low frequencies, It has good frequency and poor time resolution. Both time and frequency analysis of signals can be done simultaneously using wavelets the energy of wavelets is concentrated in terms of time and still possesses the wave-like (periodic) characteristics. So far wavelet transform has become a unified framework for denoising due to its various features. In next two subsections the decomposition process and basis function which is also called wavelet families are discussed.

A. Decomposition Process

Initially, the image is low and high pass filtered along the rows and the outcomes of each filter are down-sampled by two. Those two sub-signals represent the high frequency component and low frequency components along the rows respectively and each of size N X N/2. Then each of these sub-signals is again high and low-pass filtered, along the columns. The outcomes obtained are once more down-sampled by two.

![Fig.1. Wavelet decomposition](Image 460x526 to 540x617)

As a result the original image is divided into four sub-images each of size N/2 X N/2 containing information from different frequency components. The LL sub band is the result of low-pass filtering both the rows and columns and it contains a rough description of the image as such. Hence, the LL sub band is also known as the approximation sub band. The HH sub band is obtained by high-pass filtering in both directions and includes the high-frequency components along the diagonals as well. The HL and LH sub bands are the outcome of low-pass filtering along rows and high-pass filtering along column. LH sub band has typically the vertical detail information. HL represents the horizontal detail information. Decomposition steps are shown in fig. 1. All three sub bands HL, LH and HH are also known as the detail sub bands, because they append the high-frequency information or details to the approximation sub band.

![Wavelet Transform](Image 57x38 to 289x221)

Fig.2. Wavelet transform of 1 level on image Lena

There are two ways of wavelet decomposition pyramidal and packet decomposition defined as follows-

1. **Pyramidal Decomposition**

Pyramidal decomposition is the simplest and most common form of decomposition used. In the pyramidal decomposition, further decompositions are applied to the LL sub band only.

![Fig.3. Three decomposition steps using pyramidal decomposition](Image 324x78 to 521x194)

Fig. 3 shows a diagram of three decomposition steps. At each level the detail sub bands are the final outcome and only the approximation sub band is only further decomposed.

![Fig.4. Sub band structure after two level packet decomposition.](Image 208x123 to 306x220)

- LL1LL2, LH1LL2, HL1LL2, HH1LL2
- LL1LH2, LH1LH2, HL1LH2, HH1LH2
- LL1HL2, LH1HL2, HL1HL2, HH1HL2
- LL1HH2, LH1HH2, HL1HH2, HH1HH2

Copyright © 2014 MECS

2. Packet Decomposition

For the wavelet packet decomposition, the decomposition is not only done to the approximation subband but also further wavelet decomposition of all subbands on all levels is considered. Subband structure obtained by two level packet decomposition is shown in Fig. 4.

B. Wavelet Families

There are various basis functions of wavelets. Daubechies wavelets are the most admired wavelets. They correspond to the fundamentals of wavelet signal processing and are used in various applications. These are also known as Maxflat wavelets because their frequency responses have greatest flatness at frequencies 0 and R. This is a very advantageous property in some applications. The Haar, Daubechies, Symlets and Coiflets supported orthogonal wavelets compactly [1]. Haar wavelet is one of the oldest as well as simplest wavelet that’s why; Haar wavelet is used in experimental analysis. Wavelet transform is computationally efficient, as its basis function is compact in nature [17].

III. WAVELET BASED DENOISING

In this section, the structure of denoising process is described and then various thresholding function and values are discussed.

A. Block Diagram of Wavelet Denoising

The basic block diagram of wavelet based image denoising using thresholding is shown in Fig. 5. It involves three basic steps- first step involves computation the wavelet transform of noisy image, second step is used to apply thresholding on noisy wavelet coefficient according to some rule and finally computing inverse wavelet transform of modified wavelet coefficients.

As noise has a fine grained structure in the image therefore most of the noise represented by wavelet coefficient is also at finer scale. At fine scale wavelet coefficients also carries edge information so wavelet coefficient below the value of threshold are set to zero. Edge related coefficients are usually higher than threshold. The inverse wavelet transform of modified wavelet coefficient is taken, which represents the denoised image.

B. Wavelet Shrinkage

Threshold process or wavelet shrinkage is main process responsible for denoising which depends on threshold selection and thresholding method. All denoising algorithm first finds optimum threshold value. This threshold value can applied in three ways shown in Fig. 6. Ones threshold value and way of thresholding is chosen, then thresholding function or way of shrinkage is to be chosen.

\[
\text{sgn}(x) = \begin{cases} 
1 & \text{if } x > 0 \\
0 & \text{if } x = 0 \\
-1 & \text{if } x < 0 
\end{cases}
\]

Where \( \hat{x}_{W}(t) \) represents denoised wavelet coefficients, \( W_{y}(t) \) denotes noisy wavelet coefficients, \( T \) denotes threshold value. While in Soft thresholding, coefficients less than threshold \( T \) are set to zero while important coefficients reduced by absolute threshold value [8] as represented by (2).

\[
\hat{W}_{y}(t) = \begin{cases} 
\text{sgn}(W_{y}(t)) (W_{y}(t) - T), & \text{for } |W_{y}(t)| \geq T \\
0, & \text{otherwise}
\end{cases}
\]

Where \( \text{sgn}(\cdot) \) is Signum function, which returns 1 if the element is greater than 0, 0 if it equals zero and \(-1\) if it is less than 0. \( W_{y}(t) \) represents denoised wavelet coefficients.
coefficients, \( W_j(t) \) denotes noisy wavelet coefficients, \( T \) denotes threshold value.

It has been observed that soft thresholding outperforms hard thresholding but there are some cases where hard thresholding yields a much superior result [10]. All threshold functions can further be enhanced as adaptive method to improve the result based on application, one such approaches are given in [7, 9]. Bruce and Gao have shown that hard thresholding has smaller bias but larger variance than soft thresholding [16]. They have also observed that smaller threshold value should be used for soft thresholding [16].

2. Thresholding value

Three categories of threshold selection method are Universal threshold, Sub band adaptive threshold and spatially adaptive threshold. In universal threshold a threshold value is uniquely chosen for all wavelet coefficients [3]. In Sub band adaptive thresholding, threshold value is selected differently for each detail sub band [2]. In spatially adaptive threshold, each detail wavelet coefficient has its own threshold value [4].

IV. WAVELET DENOISING METHODS

There are various image shrinkage methods based on wavelet denoising. Some popular threshold selection methods for image denoising are visu Shrink, sure Shrink, bayes Shrink, Prob Shrink, Block Shrink and Neigh Shrink Sure. The methods are described as following in chronological way-

A. Visu Shrink

This is the first attempt of introducing wavelet based denoising. Visu Shrink thresholding is done by applying universal threshold proposed in [2]. It uses the hard thresholding rule. Threshold value \( t \) is directly proportional to the noise’s standard deviation. With additive Gaussian noise assumption Visu Shrink exhibits better denoising performance than the universal threshold but Visu Shrink does not deal with minimizing the mean squared error. Threshold of visu shrink [2] is defined using following (3)-

\[
t = \hat{\sigma}_x \sqrt{2 \times \log(n)}
\]

Where, \( n \) is the size of image. \( \hat{\sigma}_x \) is standard deviation of the noise which is obtained from [2] following (4)-

\[
\hat{\sigma}_x = \frac{\text{median}(W_{ij})}{0.6745}
\]

Where, \( w_i \) is wavelet coefficient of HH band of Noisy Image.

B. Sure Shrink

Sure Shrink works on the principle of Stein’s Unbiased Risk Estimator (SURE) proposed by [3]. Threshold value \( t_j \) for each resolution level \( j \) in the wavelet transform is used, which is referred to level dependent thresholding. The Sure Shrink threshold \( t^* \) is defined as follows [3]-

\[
t^* = \min(t, \hat{\sigma}_n \sqrt{2 \times \log(n)})
\]

Where, \( t \) denotes the value that minimizes Stein’s Unbiased Risk Estimator. Sure Shrink minimizes the mean squared error and also it is smoothness adaptive, which means that if any unknown function includes abrupt changes or boundaries in the image, the reconstructed image also has the same. But in situations of extreme sparsity of the wavelet coefficients the noise contributed to the SURE profile by 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.

C. Bayes Shrink

Bayes Shrink was proposed in [5]. As noise is additive in nature so noisy image is additive sum of original image and noise, in terms of variance it can be stated that [5]-

\[
\hat{\sigma}_j^2 = \hat{\sigma}_x^2 + \hat{\sigma}_n^2
\]

Where \( \hat{\sigma}_j^2 \) is variance of noisy image \( \hat{\sigma}_x^2 \) is variance of original image and \( \hat{\sigma}_n^2 \) is variance of noise. A good estimated threshold is Bayesian threshold \( t_B \) which is defined as [5]-

\[
t_B = \frac{\hat{\sigma}_n^2}{\hat{\sigma}_x}
\]

Where \( \hat{\sigma}_x \) is obtained from following equation [5]-

\[
\hat{\sigma}_x = \sqrt{\max(\hat{\sigma}_j^2 - \hat{\sigma}_n^2, 0)}
\]

In Bayes shrink, thresholding is done at each sub band in the wavelet decomposition which improves outcome and also completely denoise the flat regions of the image. But it is less sensitive to the noise around edges.

D. Prob Shrink

This method uses a probabilistic shrinkage function [11]. It works on estimation of the probability that a given coefficient contains a significant noise-free component, which is called “signal of interest”. Then the wavelet coefficient is multiplied with the probability. The probability is calculated by using laplacian prior for noise free sub band data and additive white Gaussian noise. Prob shrink defines a signal of interest as a noise-free coefficient component that exceeds a specific threshold \( T \) and formulate the following two hypotheses: H0: the signal of interest is absent and H1: the signal of interest is present (in a given coefficients). One can see that the
smallest coefficients are heavily shrunk towards zero while the largest ones tend to remain unchanged.

E. Block Shrink

Block Shrink is a completely data-driven block thresholding approach and is also easy to implement [13]. It can decide the optimal block size and threshold for every wavelet sub band by minimizing Stein’s unbiased risk estimate (SURE). It also limits the block size search range by following [13]-

\[ 1 \leq L \leq \left( \frac{N}{2^k} \right)^{\frac{1}{2k}} \]

Table 1. PSNR of various wavelet thresholding techniques for LENA

<table>
<thead>
<tr>
<th></th>
<th>VISU HARD</th>
<th>VISU SOFT</th>
<th>SURE SOFT</th>
<th>BAYES SHRINK</th>
<th>PROB SHRINK</th>
<th>BLOCK SHRINK</th>
<th>NEIGH SHRINK SURE</th>
<th>MEDIAN FILTER</th>
<th>WIENER FILTER</th>
</tr>
</thead>
<tbody>
<tr>
<td>σ=10</td>
<td>29.13</td>
<td>26.76</td>
<td>31.85</td>
<td>32.26</td>
<td>34.93</td>
<td>34.36</td>
<td>33.31</td>
<td>32.13</td>
<td>32.63</td>
</tr>
<tr>
<td>σ=20</td>
<td>26.25</td>
<td>24.24</td>
<td>27.55</td>
<td>28.70</td>
<td>32.02</td>
<td>31.10</td>
<td>29.79</td>
<td>28.39</td>
<td>30.01</td>
</tr>
<tr>
<td>σ=30</td>
<td>24.67</td>
<td>22.89</td>
<td>24.98</td>
<td>27.17</td>
<td>30.27</td>
<td>29.24</td>
<td>27.93</td>
<td>25.53</td>
<td>27.94</td>
</tr>
<tr>
<td>σ=40</td>
<td>23.66</td>
<td>21.95</td>
<td>22.93</td>
<td>26.10</td>
<td>29.01</td>
<td>27.97</td>
<td>26.70</td>
<td>23.32</td>
<td>26.13</td>
</tr>
</tbody>
</table>

Table 2. PSNR of various wavelet thresholding techniques for BARBARA

<table>
<thead>
<tr>
<th></th>
<th>VISU HARD</th>
<th>VISU SOFT</th>
<th>SURE SOFT</th>
<th>BAYES SHRINK</th>
<th>PROB SHRINK</th>
<th>BLOCK SHRINK</th>
<th>NEIGH SHRINK SURE</th>
<th>MEDIAN FILTER</th>
<th>WIENER FILTER</th>
</tr>
</thead>
<tbody>
<tr>
<td>σ=20</td>
<td>22.86</td>
<td>21.50</td>
<td>26.12</td>
<td>25.95</td>
<td>29.20</td>
<td>28.72</td>
<td>27.47</td>
<td>23.84</td>
<td>26.31</td>
</tr>
<tr>
<td>σ=30</td>
<td>21.68</td>
<td>20.53</td>
<td>23.74</td>
<td>24.18</td>
<td>26.86</td>
<td>26.64</td>
<td>25.38</td>
<td>22.60</td>
<td>24.79</td>
</tr>
<tr>
<td>σ=40</td>
<td>21.01</td>
<td>19.87</td>
<td>22.03</td>
<td>23.03</td>
<td>25.40</td>
<td>25.29</td>
<td>24.09</td>
<td>21.36</td>
<td>23.57</td>
</tr>
</tbody>
</table>

Where, L represents block size, N is an integer with power of two, k represents the scale.

![Fig. 2 x2 block partition for a wavelet sub band](image)

The block thresholding simultaneously keeps or kills all the coefficients in groups rather than individually, enjoys a number of advantages over the conventional term-by-term thresholding. Fig. 7 represents 2 X 2 blocks which can be generalized as b_1 X b_2 blocks. The block thresholding increases the estimation precision by utilizing the dependencies of the neighbor wavelet coefficients. Block Shrink enjoys the advantages of Sure Shrink and Neigh Shrink and gets rid of their separate weaknesses.

F. Neigh Shrink Sure

In Neigh Shrink [12], for each noisy wavelet coefficient to be shrunk, a square neighboring window centered at it. In sub band thresholding, the threshold and neighboring window size keep unchanged in all sub bands. Neigh Shrink Sure [14] is an improvement over Neigh Shrink [12], which has disadvantage of using a non-optimal universal threshold value and the same neighboring window size in all wavelet sub bands. Neigh Shrink Sure. It can determine an optimal threshold and neighboring window size for every sub band by the Stein’s unbiased risk estimate (SURE) [14]. They combine the unknown noiseless coefficients from sub bands into the corresponding 1-D vector. As using stein’s approach for almost any fixed estimator based on the data, the expected loss (i.e. risk) can be estimated in an unbiased way. They estimated optimal threshold value and sub band level using following [14]-

\[ \hat{\lambda}^*, \hat{L}^* = \arg \min SURE \left( w_i, \lambda, L \right) \]

Where \( \hat{\lambda}^* \) represents optimal threshold value for sub band S, \( \hat{L}^* \) represents optimal window size for sub band S which minimizes SURE (\( W_n, \lambda, L \)).

Visu shrink introduces wavelet based denoising using universal thresholded. Sure shrink is first attempt of introducing sub band level thresholding. Bayes shrink introduces Gaussian modeling on wavelet coefficients. Prob shrink utilizes a probabilistic approach. Block shrink and neigh shrink introduce sure level approach in a block or neighborhood. These all methods have novel approaches; we try to compare experimentally which approach has an edge over other in next section.

V. EXPERIMENT AND RESULTS

Experiments are conducted at system configuration of PC with Intel®Core i5 processor, 4 GB system memory,
MATLAB 2009b as simulation software. Test images used in experiment are Lena (512X512) and Barbara (512X512). We have considered large range of noise variance so that generalized results could be obtained. Test images are corrupted by additive Gaussian noise of different standard deviation 10, 20, 30, 40. For denoising process Haar wavelet is used at decomposition level 5. Experimental results are measured by the Peak Signal-to-Noise Ratio (PSNR) in decibels (dB), which [10] is defined as:

$$\text{PSNR} = 10 \times \log_{10} \left( \frac{255^2}{\text{MSE}} \right)$$  \hspace{1cm} (11)$$

Where MSE refers to mean squared error, which is defined as follows [10]-

$$\text{MSE} = \frac{1}{N \times N} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (x_{ij} - \hat{x}_{ij})^2$$  \hspace{1cm} (12)$$

Where N x N represents size of square image, $x_{ij}$ is pixel value of original image. $\hat{x}_{ij}$ is pixel values of denoised image. The PSNR is measured in terms of bits per sample or bits per pixel. The greater PSNR value is, the better the image quality and noise suppression.

Table 1 and Table 2 show the PSNR performance of Visu Shrink, Sure Shrink, Bayes Shrink, Prob Shrink, Block Shrink and Neigh Shrink Sure on test images Lena and Barbara respectively. For both test images visu shrink has least performance as theoretically it uses universal threshold for all sub bands which is not optimal. In case of visu shrink, we have applied two thresholding methods soft and hard both and hard thresholding exhibits better performance than soft thresholding. Sure shrink performs better than Visu shrink as it introduces sub band adaptive threshold. PSNR value obtained by Visu shrink and sure shrink decreases rapidly as noise variance is increased. Bayes shrink considers Gaussian modeling so it produces better result as compared to visu shrink and sure shrink both. Prob shrink considers the probabilistic approach to shrinkage as it first determines signal of interest. In our experiment, prob shrink has obtained highest PSNR value. Block shrink uses block level approach and utilizes dependencies between coefficients.

Neigh shrink sure also utilizes neighborhood approach and SURE based shrinkage. That’s why recently proposed wavelet based method like prob shrink, block shrink and neigh shrink sure have obtained higher value of PSNR than other wavelet based methods.

Apart from wavelet domain methods in this paper, two spatial domain methods are also implemented. These two methods are median filtering and wiener filtering. Median filtering considers a 3 X 3 mask and chooses median value of mask applied to image. Wiener filtering utilizes inverse approach and most classic approach to image restoration. Median and wiener filtering has obtained better PSNR than initial wavelet based method (visu shrink and sure shrink). But it has obtained far less PSNR as compared to later proposed wavelet based methods. Hence it could be concluded that wavelet domain processing has better capability of noise reduction as compared to spatial domain processing.

Fig.8 and Fig.9 show the PSNR curve of Visu Shrink, Sure Shrink, Bayes Shrink, Prob Shrink, Block Shrink and Neigh Shrink Sure on test images Lena and Barbara respectively. From the visual results shown in fig. it can be deducted that recently proposed wavelet methods like prob shrink, block shrink and neigh shrink sure have produced better visual images.
VI. CONCLUSION

Image denoising has been a classical problem in image processing. Wavelet thresholding has become very popular way of solving this problem due its various abilities. Threshold selection is a big challenge for image denoising. This paper reviews a set of different method of wavelet denoising using thresholding which covers Visu Shrink, Sure Shrink, Bayes Shrink, Prob Shrink, Block Shrink and Neigh Shrink Sure. In this paper two spatial domain methods median and wiener filtering is also considered. Experimental analysis shows that wavelet domain method outperforms spatial domain methods. In wavelet domain, Bayes Shrink performs better that Visu Soft and Sure Soft. Recent methods like Prob Shrink, Block Shrink and Neigh Shrink Sure outperform earlier wavelet domain methods. To develop better denoising algorithm the aspects that should be considered are exploiting neighbor dependencies, spatially adaptive approach and better estimation of variance.

REFERENCES


A Comparative Study of Wavelet Thresholding for Image Denoising


Authors’ Profiles

Arun Dixit is born in India in 1990. He is currently M.Tech scholar in computer science and engineering department at Madhav Institute of Technology and Science, Gwalior, India. Prior to it, he was lecturer in Apollo Institute of Technology, kanpur, India. He has completed B.Tech in computer science and engineering from Maharana Pratap Engineering College, Kanpur, India in 2011. His research area includes Wavelet processing, Image denoising, Image enhancement and Meta-heuristic optimization.

Poonam Sharma is presently Assistant Professor in computer science and engineering department at Madhav Institute of Technology & Science, Gwalior, India. She has got PhD degree from Maulana Azad National Institute of Technology, Bhopal, India in 2001. Her research area includes Biometrics, Image processing, soft computing and pattern recognition.