

# Comparing Wavelet and Wavelet Packet Image Denoising Using Thresholding Techniques

Tapan Kumar Hazra<sup>1</sup>, Rajib Guhathakurta<sup>2</sup>

<sup>1</sup>Department of Information Technology, Institute of Engineering and Management, Salt Lake, Kolkata-700091

<sup>2</sup>Department of Information Technology, Institute of Engineering and Management, Salt Lake, Kolkata-700091

**Abstract:** *Image denoising has remained a fundamental problem in the field of image processing. With Wavelet transforms, various algorithms for denoising in wavelet domain were introduced. Wavelets gave a superior performance in image de-noising because here multi-resolution analysis is possible. Wavelet thresholding is a signal estimation technique that exploits the capabilities of wavelet transform for signal denoising. The aim of this project was to study various denoising techniques using wavelet and wavelet packets and compare them to determine the better one for image denoising. Performance of denoising algorithm is measured using quantitative performance measures such as Signal-to-Noise Ratio (SNR) and Mean Square Error (MSE).*

**Keywords:** Multi-resolution, Sub-Band Coding, Discrete Wavelet Transform, Continuous Wavelet Transform, Wavelet Packet

## 1. Introduction

In many applications, image denoising is used to produce good estimates of the original image from noisy observations. The restored image should contain less noise than the observations while still keep sharp transitions (i.e. edges). Wavelet transform, due to its excellent localization property, has rapidly become an indispensable signal and image processing tool for a variety of applications, including compression and denoising. Wavelet denoising attempts to remove the noise present in the signal while preserving the signal characteristics, regardless of its frequency content. It involves three steps: a linear forward wavelet transform, nonlinear thresholding step and a linear inverse wavelet transform. Wavelet thresholding [1, 2] (first proposed by Donoho is a signal estimation technique that exploits the capabilities of wavelet transform for signal denoising. It removes noise by killing coefficients that are insignificant relative to some threshold, and turns out to be simple and effective, depends heavily on the choice of a thresholding parameter and the choice of this threshold determines, to a great extent the efficacy of denoising. Researchers have developed various techniques for choosing denoising parameters and so far there is no "best" universal threshold determination technique. The aim of this project was to study various thresholding techniques such as SureShrink, VisuShrink and BayesShrink and determine the best one for imagedenoising.

## 2. Motivation

Wavelet theory is one of the most modern areas of mathematics. Masterfully developed by French researchers, such as Yves Meyer, Stéphane Mallat and Albert Cohen, this theory, is now used as an analytical tool in most areas of technical research: mechanical, electronics, communications, computers, biology and medicine, astronomy and so on. In the field of signal and image processing, the main applications of wavelet theory are compression and denoising. In the context of denoising, the success of techniques based on the wavelet theory is ensured by the ability of decorrelation (separation

of noise and useful signal) of the different discrete wavelet transforms. Because the signal is contained in a small number of coefficients of such a transform, all other coefficients essentially contain noise. By filtering these coefficients, most of the noise is eliminated. Thus, each method of image denoising based on the use of wavelets follows the classic method, in three steps: computing a discrete wavelet transform of the image to be denoised, filtering in the wavelet domain and the computation of the corresponding inverse wavelet transform. Throughout recent years, many wavelet transforms (WT) have been used to operate denoising. The first one was the discrete wavelet transform; it has three main disadvantages, lack of shift invariance, lack of symmetry of the mother wavelet and poor directional selectivity. These disadvantages can be diminished using a complex wavelet transform. More than 20 years ago, Grossman and Morlet developed the continuous wavelet transform. A revival of interest in later years has occurred in both signal processing and statistics for the use of complex wavelets and complex analytic wavelets, particularly in it may be linked to the development of complex-valued discrete wavelet filters and the clever dual filter bank. The complex WT has been shown to provide a powerful tool in signal and image analysis.

## 3. Denoising Procedure

The procedure to denoise an image is given as follows:

De-noised image =  $W^{-1} [T\{W(\text{Original Image} + \text{Noise})\}]$

**Step 1:** Apply forward wavelet transform to a noisy image to get decomposed image.

**Step 2:** Apply non-linear thresholding to decomposed image to remove noise.

**Step 3:** Apply inverse wavelet transform to threshold image to get a denoised image in spatial domain.

## 4. Wavelet Transform

Wavelet transform is a relatively new concept (about 20 – 25 years old). Mathematical Transformations are applied

to signals to obtain further information from that signal that is not readily available in the raw signal. Here raw signal means „Time – domain signal“.

Most of the signals in practice are time domain signal. That is, whatever the signal is measuring is a function of time. When we plot signal, one of the axis is time (independent variable) and other is usually amplitude (dependent variable). In many cases the most distinguished information is hidden in the frequency content of the signal. The frequency spectrum of a signal is basically the frequency components of that signal.

How do we measure frequency or how do find the frequency content of a signal? Answer is, Fourier Transform (the most popular Transformation technique).

If Fourier Transform of a signal in time domain is taken, the Frequency – Amplitude graph of that signal is obtained. That is now we can plot one axis being “Frequency” and other being “Amplitude”

This plot tells us how much of each frequency exists in our signal Although Fourier Transform is the probably the most popular transform being used, there are many other transform available: Hilbert Transform, Short Time Fourier Transform (STFT), Wigner Distribution, Radon Transform, and of course, **Wavelet Transform**. Every transformation technique has its own area of application with advantages and disadvantages, and the wavelet transform is also no exception.

The natural question that comes to mind is that is it necessary to have both the time and the frequency information at the same time?

As we will see soon, the answer depends on the particular application and the nature of the signal in hand. Recall that the FT gives the frequency information of the signal, which means that it tells us how much of each frequency exists in the signal, but it does not tell us **when in time** these frequency components exist. This information is not required when the signal is so-called **stationary**.

#### 4.1 Resolution problem

Let us pass the time-domain signal through various high-pass and low-pass filters, which filter out either high frequency or low frequency portions of the signal. This procedure is repeated, every time some portion of the signal corresponding to some frequencies being removed from the signal.

But there is an issue, called "**uncertainty principle**", which states that, we cannot exactly know **what frequency exists at what time instance**, but we can only know **what frequency bands exist at what time intervals**.

#### 4.2 Uncertainty Principle

The uncertainty principle, originally found and formulated by Heisenberg, states that, the momentum and the position of a moving particle cannot be known simultaneously.

This applies to our subject as follows:

Higher frequencies are better resolved in time, and lower frequencies are better resolved in frequency. This means that, a certain high frequency component can be located better in time (with less relative error) than a low frequency component. On the contrary, a low frequency component can be located better in frequency compared to high frequency component.

This leads to a Resolution Problem.

#### 4.3 Multi resolution Analysis

Although the time and frequency resolution problems are results of a physical phenomenon (the Heisenberg uncertainty principle) and exist regardless of the transform used, it is possible to analyze any signal by using an alternative approach called the multi-resolution analysis (MRA). MRA, as implied by its name, analyzes the signal at different frequencies with different resolutions. Every spectral component is not resolved equally as was the case in the STFT (Short Term Fourier Transform).

MRA is designed to give good time resolution and poor frequency resolution at high frequencies and good frequency resolution and poor time resolution at low frequencies.

This approach makes sense especially when the signal at hand has high frequency components for short durations and low frequency components for long durations.

#### 4.4 Continuous Wavelet Transform

The continuous wavelet transform was developed as an alternative approach to the short time Fourier transforms to overcome the resolution problem. The wavelet analysis is done in a similar way to the STFT analysis, in the sense that the signal is multiplied with a function, similar to the window function in the STFT, and the transform is computed separately for different segments of the time-domain signal.

However, there are two main differences between the STFT and the CWT:

1. The Fourier transforms of the windowed signals are not taken, and therefore single peak will be seen corresponding to a sinusoid, i.e., negative frequencies are not computed.
2. The width of the window is changed as the transform is computed for every single spectral component, which is probably the most significant characteristic of the wavelet transform.

The continuous wavelet transform is defined as follows

$$CWT_x^{\phi}(\tau, S) = \psi_x^{\phi}(\tau, S) = \frac{1}{\sqrt{|S|}} \int x(t) \psi^* \left( \frac{t-\tau}{S} \right) dt$$

As seen in the above equation, the transformed signal is a function of two variables, **tau** and **s**, the **translation** and **scale** parameters, respectively. **psi(t)** is the transforming function, and it is called **the mother wavelet**. The term **mother wavelet** gets its name due to two important properties of the wavelet analysis as explained below:

The term **wavelet** means a **small wave**. The smallness refers to the condition that this (window) function is of finite length (**compactly supported**). The wave refers to the condition that this function is oscillatory. The term **mother** implies that the functions with different region of support that are used in the transformation process are derived from one main function, or the mother wavelet. In other words, the mother wavelet is a **prototype** for generating the other window functions.

The term **translation** is used in the same sense as it was used in the STFT; it is related to the location of the window, as the window is shifted through the signal. This term, obviously, corresponds to time information in the transform domain. However, we do not have a frequency parameter, as we had before for the STFT. Instead, we have scale parameter. The term frequency is reserved for the STFT.

#### 4.5 Discrete Wavelet Transform (DWT)

The DWT of an image  $x$  is calculated by passing it through a series of filters. First the samples are passed through a low pass filter with impulse response  $g$  resulting in a convolution of the two.

The image is also decomposed simultaneously using a high-pass filter  $h$ . The outputs give the detail coefficients (from the high-pass filter) and approximation coefficients (from the low-pass filter). It is important that the two filters are related to each other and they are known as a quadrature mirror filter. However, since half the frequencies of the signal have now been removed, half the samples can be discarded according to Nyquist's rule. The filter outputs are then down sampled by 2.

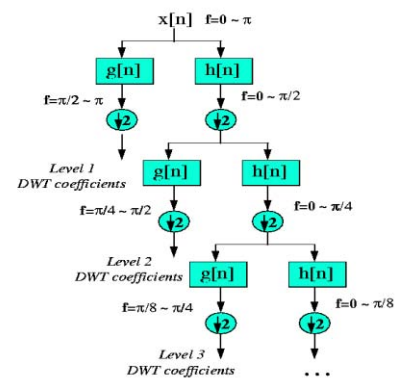
### 5. Sub-Band Coding

Sub-band coding [9] is a method for calculating the Discrete Wavelet Transform. The whole sub-band process consists of a filter bank, and filters of different cutoff frequencies are used to analyze the signal at different scales.

The procedure starts by passing the signal through a half band high-pass filter and a half band low-pass filter. A half band low-pass filter eliminates exactly half the frequencies from the low end of the frequency scale. For example, if a signal has a maximum of 1000 Hz component, then half band low-pass filter removes all the frequencies above 500 Hz. The filtered signal is then down sampled; meaning

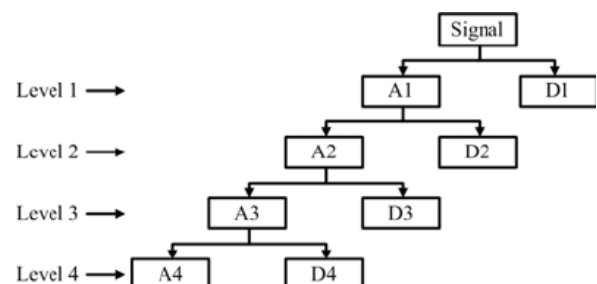
some sample of the signal is removed. Then the resultant signal from the down sampled half band low-pass filter is then processed in the same way again. This process will produce sets of wavelet transform coefficients that can be used to reconstruct the signal. The resolution of the signal is changed by filtering operations, and the scale is changed by down sampling operations. Down sampling a signal corresponds to reducing the sampling rate, which is equivalent to removing some of the samples of the signal.

The figure 1 illustrates this decomposition procedure, where  $x[n]$  is the original signal to be decomposed, and  $h[n]$  and  $g[n]$  are low-pass and high-pass filters, respectively. The bandwidth of the signal at every level is marked on the figure as "f"



**Figure 1:** Wavelet Decomposition

The frequencies that are most prominent in the original signal will appear as high amplitudes in that region of the DWT signal that includes those particular frequencies. The difference of this transform from the Fourier transform is that the time localization of these frequencies will not be lost. However, the time localization will have a resolution that depends on which level they appear. If the main information of the signal lies in the high frequencies, as happens most often, the time localization of these frequencies will be more precise, since they are characterized by more number of samples. If the main information lies only at very low frequencies, the time localization will not be very precise, since few samples are used to express signal at these frequencies. This procedure in effect offers a good time resolution at high frequencies, and good frequency resolution at low frequencies. Most practical signals encountered are of this type.

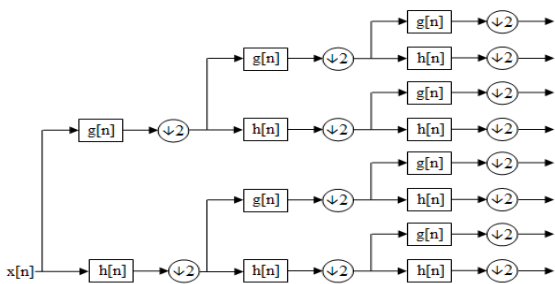


**Figure 2:** Level 4 decomposition using wavelet transform

## 6. Wavelet Packet Decomposition

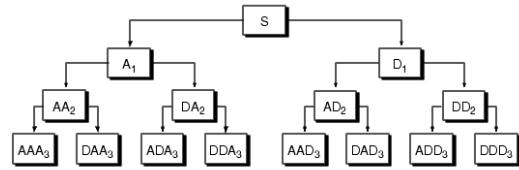
Wavelets are very effective for coding locally smooth images. However, wavelets are not very efficient in representing texture-rich images like finger print images and remote sensing images. The reason is that, texture-rich images are mainly described by smaller scale wavelet coefficients. These smaller scale coefficients carry very little energy, and are often quantised to zero, even at high bit rate. Wavelet packets with much larger libraries of functions can more efficiently represent texture-rich images than wavelets.

The wavelet packet method is a generalization of wavelet decomposition that offers a richer range of possibilities for signal analysis and which allows the best matched analysis to a signal. It provides level by level transformation of a signal from the time domain into the frequency domain. It is calculated using a recursion of filter-decimation operations leading to the decrease in time resolution and increase in frequency resolution. The frequency bins, unlike in wavelet transform, are of equal width, since the WPT divides not only the low, but also the high frequency subband. In wavelet analysis, a signal is split into an approximation and a detail coefficient. The approximation coefficient is then itself split into a second-level approximation coefficients and detail coefficients, and the process is repeated. In wavelet packet analysis, the details as well as the approximations can be split. This yields more than different ways to en-code the signal. When the WT is generalized to the WPT, not only can the low-pass filter output be iterated through further filtering, but the high-pass filter can be iterated as well. This ability to iterate the high-pass filter outputs means that the WPT allows for more than one basis function (or *wavelet packet*) at a given scale, versus the WT which has one basis function at each scale other than the deepest level, where it has two. The set of *wavelet packets* collectively make up the complete family of possible bases, and many potential bases can be constructed from them. If only the low-pass filter is iterated, the result is the wavelet basis. If all low-pass and high-pass filters are iterated, the *complete tree* basis results. The top level of the WPD tree is the time representation of the signal. As each level of the tree is traversed there is an increase in the trade-off between time and frequency resolution. The bottom level of a fully decomposed tree is the frequency representation of the signal. The following figure shows the level 3 decomposition using wavelet packet transform.



**Figure 3:** Wavelet Packet decomposition over 3 levels.  $g[n]$  is the low-pass approximation coefficients,  $h[n]$  is the high-pass detail coefficients

For  $n$  levels of decomposition the WPD produces  $2^n$  different sets of coefficients (or nodes) as opposed to  $(3n + 1)$  sets for the DWT. However, due to the down sampling process the overall number of coefficients is still the same and there is no redundancy.



**Figure 4:** Level 3 decomposition using wavelet packet transform

Based on the above analysis, give the comparison of a three-level wavelet decomposition and wavelet packet decomposition. It can be seen in **Figure 1** that in wavelet analysis only the approximations (represented by capital A in the figure) at each resolution level are decomposed to yield approximation and detail information (represented by capital D in the figure) at a higher level. However, in the wavelet packet analysis [Figure 4], both the approximation and details at a certain level are further decomposed into the next level, which means the wavelet packet analysis can provide a more precise frequency resolution than the wavelet analysis.

## 7. Thresholding Techniques

Thresholding is a simple non-linear technique, which operates on one wavelet coefficient at a time. In its most basic form, each coefficient which is smaller than threshold, set to zero, otherwise, it is kept or modified. The small co-efficient are dominated by noise, while coefficient with large absolute value carry more signal information than noise. Replacing noise co-efficient (small coefficients below a certain threshold value) by zero and an inverse wavelet transform may lead to a reconstruction that has lesser noise. This thresholding idea is based on the following:

1. The de-correlating property of wavelet transform creates a sparse signal. Most untouched coefficient is zero or close to zero.
2. Noise is spread out equally along all co-efficient.
3. The noise level is not too high so that one can distinguish the signal wavelet coefficients from binary ones.

This method is an effective and thresholding is simple and efficient method for noise reduction.

### 7.1 Threshold selection

As one may observe, threshold selection is an important question when denoising. A small threshold may yield a result close to the input, but the result may still be noisy. A large threshold on the other hand, produces a signal with a large number of zero coefficients. This leads to a smooth signal. Paying too much attention to smoothness, however, destroys details and in image processing may cause blur and blocking artefacts.

### 7.2 Hard Thresholding

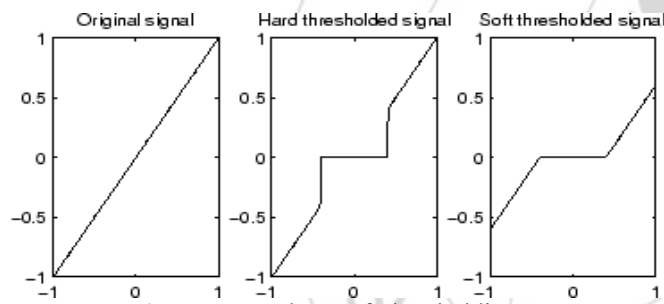
One of the most attractive features of wavelet thresholding is that, for the type of random noise frequently encountered, in signal transmission, it is possible to automatically choose a threshold for denoising without any prior knowledge of the signal.

By choosing a threshold that is significantly large, and multiplying with the standard deviation of the random noise, it is possible to remove most of the noise by thresholding the wavelets transform coefficients. This process is known as *hard* threshold

$$T_{\tau}^{hard} = \begin{cases} Y[m, n], & |Y[m, n]| > \tau \\ 0, & |Y[m, n]| \leq \tau \end{cases}$$

where,  $\tau$  is the threshold value.

From Figure 5, we can see that hard thresholding can create discontinuities, and thus greatly exaggerates small differences in the transform value whose magnitudes are near the threshold value  $\tau$ . If the value is only slightly less than  $\tau$ , then this value is set equal to zero, while a value whose magnitude is only slightly greater than  $\tau$  is left unchanged. Therefore, hard thresholding is not suitable for most noise removal.



**Figure 5:** Hard vs. soft thresholding

### 7.3 Soft Thresholding

With a slight modification to the hard thresholding approach, a method known as *soft* thresholding can be created

$$T_{\tau}^{soft} = \begin{cases} \text{sgn}(Y[m, n]) (|Y[m, n]| - \tau), & |Y[m, n]| > \tau \\ 0, & |Y[m, n]| \leq \tau \end{cases}$$

Where,  $\tau$  is the threshold value.

### 7.4 SURE Shrink

SureShrink is a thresholding by applying subband adaptive threshold [4], a separate threshold is computed for each detail subband based upon SURE (Stein's unbiased estimator for risk), a method for estimating the

loss  $\|\hat{A} - \mu\|$  in an unbiased fashion. In our case let wavelet coefficients in the  $j$ th subband be  $\{X_i : i = 1, \dots, d\}$ ,  $\hat{A}$  is the soft threshold estimator. SURE Shrink is a thresholding applied to sub-band adaptively.

It is based on Stein's Unbiased Risk Estimator (SURE), a method for estimating the loss in an unbiased fashion

Let wavelet coefficients in the  $j$ th sub-band be  $\{X_i : i = 1, \dots, d\}$ ,

For the soft threshold estimator  $\hat{X}_i = \eta_{\tau}(X_i)$ ,

we have  $SURE(\tau, X) = d - 2 \sum_{i=1}^d \{i : |X_i| \leq \tau\} + \sum_{i=1}^d \min\{|X_i|, \tau\}^2$

Select threshold  $\tau^* = \arg \min_{\tau} SURE(\tau, X)$

## 8. Experimental Results



**Figure 6(a-h):** Global Thresholding using SURE Shrink with various wavelets and wavelet packets

**Fig a:** Original Gray Scale Image

**Fig b:** Noisy Image after adding Gaussian Noise ( $\sigma = 0.003$ )

**Fig c:** Denoised image using Wavelet Packet, Global Thresholding, Symlet4

**Fig d:** Denoised image using Wavelet Packet, Global Thresholding, db2 level2

**Fig e:** Denoised image using Wavelet Packet, Global Thresholding, db4 level4

**Fig f:** Denoised image using Wavelet, Global Thresholding, Symlet4

**Fig g:** Denoised image using Wavelet, Global Thresholding, db2 level2

**Fig h:** Denoised image using Wavelet, Global Thresholding, db4 level4

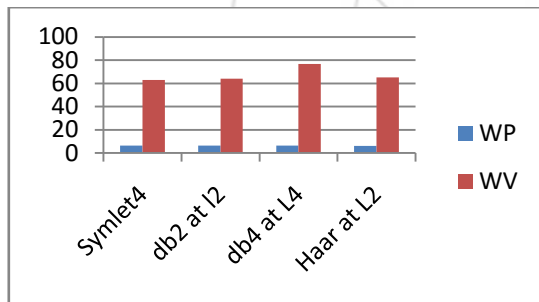
## 9. Results and Discussion

**Table 1:** Comparison of MSE & PSNR For Different  $\sigma$

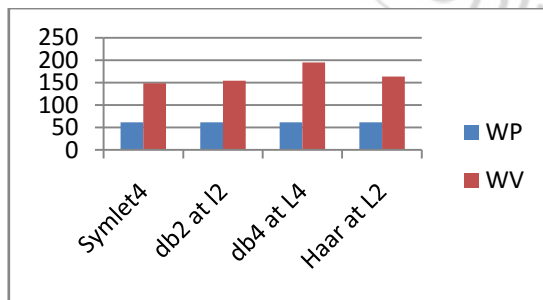
MSE					PSNR			
Description Of the wavelet Packet used								
$\sigma$	0.0001	0.001	0.01	0.1	0.0001	0.001	0.01	0.1
Symlet4	6.37	61.65	594.70	4551.06	40.13	30.27	20.42	11.58
Daubechies2 at level 2	6.28	61.56	594.73	4551.22	40.18	30.27	20.42	11.58
Daubechies4 at level 4	6.46	61.63	594.62	4551.19	40.06	30.27	20.42	11.58
Haar at level 2	6.17	61.57	594.77	4551.06	40.07	30.27	20.42	11.58
Description Of the wavelet used								
Symlet4	63.12	148.46	306.95	749.67	30.16	26.45	23.29	19.42
Daubechies2 at level 2	64.22	154.64	322.50	777.23	30.09	26.27	23.08	19.26
Daubechies4 at level 4	76.90	195.27	448.86	925.66	29.31	25.26	21.64	18.50
Haar at level 2	65.06	163.87	371.68	860.27	29.31	25.26	21.64	18.50

### 9.1 Result Analysis

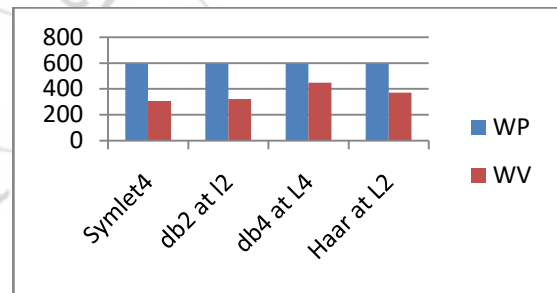
Comparison of MSE for Different  $\sigma$  Value when denoising through various wavelet Packets (WP) and Wavelets (WV) are shown in Figure 6(a-d) and that of PSNR are shown in the Figure 7(a-d).



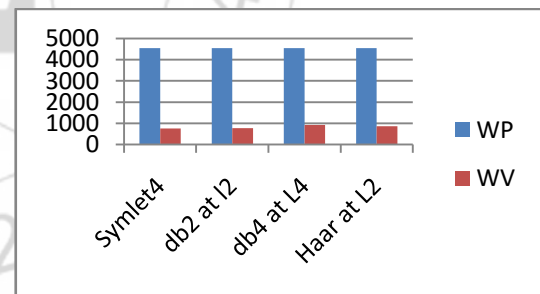
**Figure 7 (a):** MSE for  $\sigma = 0.0001$



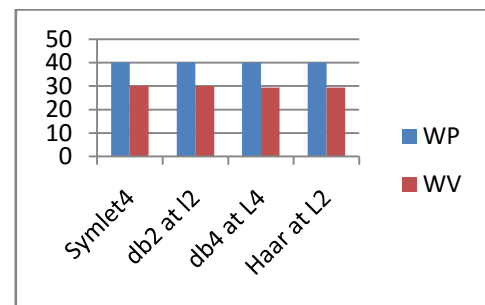
**Figure 7 (b):** MSE for  $\sigma = 0.001$



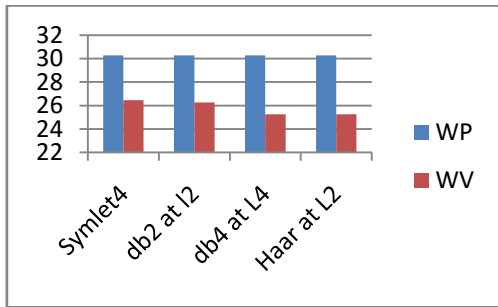
**Figure 7 (c):** MSE for  $\sigma = 0.01$



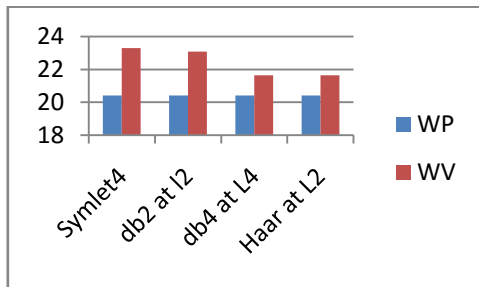
**Figure 7 (d):** MSE for  $\sigma = 0.1$



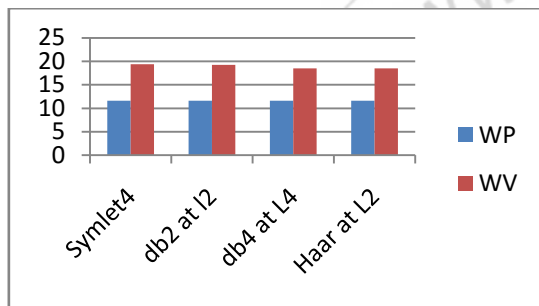
**Figure 8 (a):** PSNR for  $\sigma = 0.0001$



**Figure 8(b):** PSNR for  $\sigma = 0.001$



**Figure 8 (c):** PSNR for  $\sigma = 0.01$



**Figure 8(d):** PSNR for  $\sigma = 0.1$

## 10. Conclusions

In this work we compared some denoising methods with the SURE Shrink which uses classical wavelet and wavelet packet for denoising an image. From the obtained results it can be seen that SURE Shrink gives the better PSNR when decomposed with wavelet packet at comparatively lower noise intensity. But when noise intensity is comparatively higher, classical wavelets give better MSE & PSNR value than wavelet packet.

## References

- [1] Iain M. Johnstone David L Donoho. Adapting to smoothness via wavelet shrinkage. *Journal of the Statistical Association*, 90(432):1200–1224, Dec 1995.
- [2] David L Donoho. De-noising by soft thresholding. *IEEE Transactions on Information Theory*, 41(3):613–627, May 1995.
- [3] S. Mallat, “A Wavelet Tour of Signal Processing,” Cambridge University Press, New York, 1999.
- [4] Martin Vetterli S Grace Chang, Bin Yu. Adaptive wavelet thresholding for image denoising and compression. *IEEE Transactions on Image Processing*, 9(9):1532–1546, Sep 2000
- [5] Abdolhossein Fathi and Ahmad Reza Naghsh-Nilchi, “Efficient Image Denoising Method Based on a New

Adaptive Wavelet Packet Thresholding Function” *IEEE transaction on image processing*, vol. 21. No.9, September 2012.

- [6] S. Lee, 1980, "Digital image enhancement and noise filtering by use of local statistics," *IEEE Trans. on PAMI*, Vol.2, No.2, pp.165- 168.
- [7] D. L. Donoho and I. M. Johnstone, “Ideal Spatial Adaptation by Wavelet Shrinkage,” *Biometrika*, Vol. 81, No. 3, 1994, pp. 425-455.
- [8] I. Daubechies, “The Wavelet Transform, Time-Frequency Localization and Signal Analysis,” *IEEE Transaction on Information Theory*, Vol. 36, No. 5, 1990, pp. 961-1005.
- [9] E. P. A. Montuori, “Real time performance measures of low delay perceptual audio coding,” *Journal of Electrical Engineering*, Vol. 56, No. 3–4, pp. 100–105, 2005

## Author Profile



**Tapan Kumar Hazra** completed his M.E degree from Jadavpur University, Kolkata, West Bengal, India. Since from 2003, he is working as Assistant Professor of Department of Information Technology at Institute of Engineering & Management, Salt Lake, Kolkata, West Bengal, India. His research interest includes Design and Analysis of Algorithms, Image Processing, Natural Language Processing, Sentiment Analysis, Machine learning, Cryptography.



**Rajib Guhathakurta** is pursuing M.Tech in Information Technology from Institute of Engineering and Management (IEM), Kolkata, West Bengal, India. His research interest includes Image Processing and Networking.