International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 2 Issue 9, September - 2013

Comparison of Various Thresholding Techniques of Image Denoising

Shivani Mupparaju [1], B Naga Venkata Satya Durga Jahnavi [2]
Department of ECE, VNR VJIET, Hyderabad, A.P, India

Abstract

Denoising using wavelets attempts to remove the noise present in the signal while signal characteristics are preserved, regardless of its frequency content. It can be handled using three steps: a linear forward wavelet transform, nonlinear thresholding step and a linear inverse wavelet transform. Wavelet denoising is a lot different from smoothing; smoothing is used to remove the high frequencies and retains the lower frequencies. Wavelet shrinkage is a non-linear process and it is used to distinguish from entire linear denoising technique. Wavelet shrinkage depends on the choice of a thresholding parameter and the choice of how the threshold is determined, and the efficacy of denoising various techniques can be used for choosing denoising parameters and so far there is no "best" universal threshold determination technique. So, Various denoising techniques such as Sure Shrink, Bayes Shrink and Visu Shrink determines the best one for image denoising.

1. Introduction

During acquisition and transmission, image denoising can used to remove the additive noise while keeping the important signal features. In the recent years wavelet thresholding and threshold selection for signal de-noising has gain more interest because wavelet gives an appropriate basis for separating noisy signal from the image signal. The motivation is the wavelet transform is good at compacting energy, the small coefficient are more likely due to noise and large coefficient due to important signal features . These small coefficients can be thresholded without affecting the main features of the image.

A simple non-linear technique called 'thresholding' which operates on one wavelet coefficient at a time. Each coefficient is thresholded by comparing against threshold. And if the coefficient is smaller than the threshold, set it to zero; otherwise it is modified. Replacing the small coefficients by zero and applying inverse wavelet transform on the result may lead to reconstruction retaining back the essential signal characteristics and giving less noise image.

2. Objectives and Tools Employed

2.1. Objective of the project

The main objective of this paper is study various thresholding techniques such as Sure Shrink, Visu Shrink and Bayes Shrink and determine the best one for image denoising.

2.2. Tools Used

Software: MATLAB

3. Types of Noise

3.1. Gaussian Noise

Because of its mathematical tractability in both spatial and frequency domains, Gaussian (also called normal) noise models are used frequently in practice. In fact, this tractability is so convenient that it often results in Gaussian models being used in situations in which they are marginally applicable at best.

The Probability density function of Gaussian random variable z, is given by:

$$P(z) = \frac{1}{\sqrt{2\pi\sigma}} e^{-(z-\bar{z})/2\sigma^2} \quad (1.1)$$

where z represents intensity, $\bar{z}$ is the mean (average) value of z , and σ is its standard deviation. The standard deviation squared σ^2, is called variance of z.

3.2 Rayleigh Noise

The probability density function of Rayleigh Noise is given by

$$P(z) = \frac{2}{b}(z-a)e^{-(z-a)^2/b} \quad \text{for } z >= a$$
$$= 0 \quad \text{for } z < a \qquad (1.2)$$

The mean and variance of this density are given by

$$\bar{z} = a + \sqrt{\pi b/4} \qquad (1.3)$$

And

$$\sigma^2 = \frac{b(4-\pi)}{4} \qquad (1.4)$$

3.3. Erlang (Gamma) Noise

The probability density function of Erlang noise is given by

$$P(z) = \frac{a^b z^{b-1}}{(b-1)!}e^{-az} \quad \text{for } z >= 0$$
$$= 0 \quad \text{for } z < 0 \qquad (1.5)$$

Where the parameters are such that a > 0, b is a positive integer. The mean and variance of this density are given by

$$\bar{z} = \frac{b}{a} \qquad (1.6)$$

And

$$\sigma^2 = \frac{b}{a^2} \qquad (1.7)$$

3.4. Exponential Noise

The Probability density function of exponential noise is given by

$$P(z) = ae^{-az} \quad \text{for } z >= 0$$
$$= 0 \quad \text{for } z < 0 \qquad (1.8)$$

Where a > 0. The mean and variance of this density function are

$$\bar{z} = \frac{1}{a} \qquad (1.9)$$

And

$$\sigma^2 = \frac{1}{a^2} \qquad (1.10)$$

3.5. Uniform Noise

The Probability density function is given by

$$P(z) = \frac{1}{b-a} \quad \text{if } a <= z <= b$$
$$= 0 \quad \text{otherwise} \qquad (1.11)$$

The mean of this density function is given by

$$\bar{z} = \frac{a+b}{2} \qquad (1.12)$$

And its variance by

$$\sigma^2 = \frac{(b-a)^2}{12} \qquad (1.13)$$

4. Denoising

In many cases, additive noise is evenly distributed over the frequency domain (i.e., white noise), whereas an image contains mostly low frequency information. The noise is a characteristic at high frequencies and its effects can be reduced using low-pass filter. So a frequency filter or with a spatial filter can also be used. Often a spatial filter is preferably used, as it is computationally less expensive than a frequency filter. Denoising can be done in various domains and by using various methods

a) Spatial domain
b) Frequency domain
c) Wavelet domain and
d) Curvelet domain

5. Thresholding

5.1 Motivation for Wavelet thresholding

The plot of wavelet coefficients suggests that small coefficients are decreased due to noise, while coefficients with a large absolute value carry more signal information. Replacing noisy coefficients (small coefficients below a certain threshold value) by zero and an inverse wavelet transform may lead to a reconstruction that has lesser noise. Stated more precisely, we are motivated to this thresholding idea based on the following assumptions:

i. The deco-relating property of a wavelet transform creates a sparse signal: most coefficients, which are zero or close to zero when they are left untouched.
ii. Noise is spread out equally along all coefficients.
iii. The noise level is not too high so that we can distinguish the signal wavelet coefficients from the noisy ones.

As it turns out, this method is indeed effective and thresholding is a simple and efficient method for noise reduction. Further, inserting zeros creates more sparsity in the wavelet domain and here we see a link between wavelet de-noising and compression.

5.2 Hard and soft thresholding:

Hard and soft thresholding with threshold ¸ are defined as follows The hard thresholding operator is defined as

$D(U,\lambda) = U$ for all $|U| > \lambda$¸
$= 0$ otherwise (2.1)

The soft thresholding operator can be defined as

$D(U,\lambda) = sgn(U)max(0,U|-\lambda)$ (2.2)

Hard threshold is a "keep or kill" procedure and is more intuitively appealing. The transfer function of the same is shown. The alternative, soft thresholding (whose transfer function is shown), shrinks coefficients above the threshold in absolute value. While hard thresholding may seem goodl, the continuity of soft thresholding has some advantages. Moreover, hard thresholding does not even work with some algorithms such as the GCV procedure. At times, pure noise coefficients may pass the hard threshold and appear as annoying 'blips' in the output. Soft thresholding shrinks these false structures.

5.3 Threshold determination

Threshold determination is an important in image denoising. A small threshold gives a result close to the input, but the result can still have the noise component. Whereas a large threshold, produces a signal with a large number of zero coefficients. This results in smooth noiseless signal.

The effect of smoothness, however, destroys details and in image processing may cause blur and artifacts.

To investigate the effect of threshold selection, we performed wavelet denoising using hard and soft thresholds on four signals popular in wavelet literature: Blocks and Doppler.

The setup is as follows:

1. The original signals have length 2048.
2. We step through the thresholds from 0 to 5 with steps of 0.2 and at each step denoise the four noisy signals by both hard and soft thresholding with that threshold.
3. For each threshold, the MSE of the denoised signal is calculated.
4. Repeat the above steps for different orthogonal bases, namely, Haar, Daubechies 2,4 and 8.

Fig:Table 1

The results are tabulated in the Table 1 and represents Blocks and Doppler of both hard and soft for different filters.

5.4 Comparison with Universal threshold:

The threshold $\lambda_{UNIV} = \sqrt{2\ln N\sigma}$ (N being the signal length, σ^2being the noise variance) is well known in wavelet literature as the Universal threshold. It is the optimum threshold and minimizes the cost function of the difference between the function and the soft threshold version of the same in the L2.In our case, N=2048, $\sigma = 1$, therefore theoretically,

$$\sqrt{2\ln(2048)}(1) = 3.905$$

As seen from the table, the best empirical thresholds for both hard and soft thresholding are much lower than this value, independent of the wavelet used. It therefore seems that the universal threshold is not useful to determine a threshold. However, it is useful for obtain a starting value when nothing is known of the signal condition. One can surmise that the universal threshold may give a better estimate for the soft threshold if the numbers of samples are larger.

5.5 Image Denoising using Thresholding:

An image is often corrupted by noise in its acquisition or transmission. The underlying concept of denoising in images is similar to the 1D case. The goal is to remove the noise while retaining the important signal features as much as possible.

The noisy image is represented as a two-dimensional matrix $\{x_{ij}\}$,i ,j=1,….N. The noisy version of the image is modeled as

$y_{ij}=x_{ij}+n_{ij}$ ij=1,…….N. (2.3)

Where $\{n_{ij}\}$ are iid as N $(0,\sigma^2)$. We can use the same principles of thresholding and shrinkage to achieve denoising as in 1-D signals. The problem again boils down to finding an optimal threshold such that the mean squared error between the signal and its estimate is minimized.

The wavelet decomposition of an image is done as follows: In the first level of decomposition, the image is split into 4 sub bands, namely the HH, HL, LH and LL sub bands. The HH sub band gives the diagonal details of the image the HL sub band gives the horizontal features while the LH sub band represents the vertical structures. The LL sub band is the low-resolution residual consisting of low frequency components and it is this sub band, which is further split at higher levels of decomposition. The different methods for denoising we investigate differ only in the selection of the threshold. The basic procedure remains the same:

i. Calculate the DWT of the image.
ii. Threshold the wavelet coefficients.
iii. Compute the IDWT to get the denoised estimate.

Moreover, it is also found to yield visually more pleasing images. Hard thresholding introduces artifacts in the recovered images.
The three thresholding techniques- Visu Shrink, Sure Shrink and Bayes Shrink and investigate their performance for denoising various standard images

5.5.1 VisuShrink: Visu Shrink was introduced by Donoho . It can be defined as $\sigma\sqrt{2\log I}$ where σ is the noise variance and I is the number of pixels in the image.The maximum of any I values can be given by $N(0,\sigma^2)$ with the probability approaching 1 as number of pixels in the given image increases. Therefore if it has high probability, a pure noise signal is calculated as being identically zero.

However, for denoising images, Visu shrink is found to yield an overly smoothed estimate as seen. This is because the universal threshold (UT) is derived under the constraint that with high probability the estimate should be at least as smooth as the signal. The universal threshold is high for large values of I, killing many signal coefficients along with the noise. Thus, the threshold doesnt perform well at discontinuities in the signal.

5.5.2 Sure Shrink: Let $\mu = (\mu_i : i = 1,\ldots\ldots d)$ be a length-d vector, and let $x = \{x_i\}$ (with xi distributed as N(¹i,1)) be multivariate normal observations with mean vector μ. Let $\hat{\mu} = \hat{\mu}(x)$ be a fixed estimate based on the observations x. SURE can be defined as Stein's unbiased Risk Estimator). It is a special method for estimating the loss in an unbiased fashion where $\hat{\mu}$ is the soft threshold estimator. Stein's result to get an unbiased estimate of the risk is applied. For an observed vector x is the set of noisy wavelet coefficients in a sub band we find out the threshold t^s that minimizes SURE (t,x) ,i.e.

$$t^s = \operatorname{argmin}_t \, SURE(t,x) \qquad (2.4)$$

The above optimization problem is computationally straightforward. Without loss of generality, we can reorder x in order of increasing $|x_i|$. Then on intervals of t that lie between two values of $|x_i|$, SURE (t) is strictly increasing. Therefore the minimum value of t^s is one of the data values $|x_i|$ there are only d values and the threshold can be obtained.

5.5.3 Threshold Selection in Sparse Cases :The drawback of SURE in situations of extreme sparsity of the wavelet coefficients. In such cases the noise contributed to 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. The idea behind this hybrid scheme is that the losses while using an universal threshold, $t_d^F = \sqrt{2\log d}$ tend to be larger than SURE for dense situations, but much smaller for sparse cases .So the threshold is set to t^F_d in dense situations and to t^S in sparse situations. Thus the

$$\hat{\mu}^x(x)_i = \begin{cases} \eta_{t_d^F}(x_i) & s_d^2 \le \gamma_d \\ \eta_{t^s}(x_i) & s_d^2 > \gamma_d \end{cases} \qquad (4.12)$$

Where $s_d^2 = \frac{\sum_i (x_i^2 - 1)}{d}$

and

$$\gamma_d = \frac{\log_d^{3/2}(d)}{\sqrt{d}} \qquad (2.5)$$

η being the thresholding operator.
estimator in the hybrid method works as follows

5.5.4 SURE applied to image denoising: The wavelet decomposition of the noisy image is obtained. The SURE threshold is determined for each sub band using the above equations. We choose between this threshold and the universal threshold using the equation .The expressions s_d^2 and γ_d in the equations, given for σ = 1 have to suitably modified according to the noise variance σ and the variance of the coefficients in the sub band. The results obtained for the image 'lina' (512*512pixels) using Sure Shrink are shown in results. The `Db4' wavelet was used with 4 levels of decomposition. Clearly, the results are much better than Visu Shrink. The sharp features of the image are retained and the MSE is considerably lower. This is because Sure Shrink is sub band adaptive- a separate threshold is computed for each detail sub band.

5.5.5 Bayes Shrink: In Bayes Shrink we determine the threshold for each subband assuming a Generalized Gaussian Distribution (GGD) . The GGD is given by

$$GG_{\sigma x,\beta}(x) = C(\sigma_x,\beta)\exp - [\alpha(\sigma_x,\beta)\,|x|]^{\beta}$$

$$-\infty < x < \infty, \beta > 0 \qquad (2.6)$$

Where

$$\alpha(\sigma_x,\beta)=\alpha_X^{-1}\left[\frac{\Gamma(3/\beta}{\Gamma(1/\beta)}\right]^{1/2} \quad (2.7)$$

and $$C(\sigma_x,\beta)=\frac{\beta.\alpha(\sigma_x,\beta)}{2\Gamma(1/\beta)} \quad (2.8)$$

and $$\Gamma(t)=\int_0^{\infty} e^{-u}u^{t-1}du \quad (2.9)$$

The parameter σ_x is the standard deviation and β is the shape parameter It has been observed that with a shape parameter β ranging from 0.5 to 1, we can describe the distribution of coefficients in a sub band for a large set of natural images. Assuming such a distribution for the wavelet coefficients, we empirically estimate β and σ_x for each sub band and try to find the threshold T which minimizes the Bayesian Risk, i.e., the expected value of the mean square error.

$$\tau(T)=E(\hat{X}-X)^2=E_xE_{y/x}\left(\hat{X}-X\right)^2$$

Where $\hat{X}$ = η_T(Y); Y/X ~ N(x,σ^2) and X ~ G $G_{x,\beta}$.
The optimal threshold T* is then given by

$$T^*(\sigma_x,\bar{\beta}) = \arg\min \Gamma(T)$$

It is a function of the parameters σ_x and β. Since there is no closed form solution for T*, numerical calculation is used to find its value.
It is observed that the threshold value set by

$$T_B(\sigma_x)=\frac{\sigma^2}{\sigma_x} \quad (2.10)$$

is very close to T*. The estimated threshold TB = σ^2=σ_x is not only nearly optimal but also has an intuitive appeal. The normalized threshold, $T_{B/\sigma}$. is inversely proportional to σ_x, the standard deviation of X, and proportional to σ_x, the noise standard deviation. When σ/σ_x<<1, the signal is much stronger than the noise, $T_{b/\sigma}$ is chosen to be small in order to preserve most of the signal and remove some of the noise; when σ = σ_x>>1, the noise dominates and the normalized threshold is chosen to be large to remove the noise which has exacted the signal. Thus, this threshold choice adapts to both the signal and the noise characteristics as reflected in the parameters σ and σ_x..

5.5.6 Parameter Estimation to determine the Threshold:

The GGD parameters, σ_x and β, need to be estimated to compute T_B (σ_x) . The noise variance σ^2 is estimated from the subband HH1 by the robust median estimator ,

$$\hat{\sigma}=\frac{median\left(|Y_{ij}|\right)}{0.6745}$$

$$Y_{ij}\in subbandHH_1 \quad (2.11)$$

The parameter β does not explicitly enter into the expression of T_B (σ_x). Therefore it suffices to estimate directly the signal standard deviation σ_x. The observation gives Y = X + V, with X and V as independent of each other, hence

$$\sigma_Y^2=\sigma_X^2+\sigma^2 \quad (2.12)$$

Where σ_Y^2 is the variance of Y. Since Y is modeled as zero-mean, σ_Y^2 can be found empirically by

$$\hat{\sigma}_Y^2=\frac{1}{n}\sum_{i,j=1}^{n}Y_{i,j}^2 \quad (4.20) \quad (2.13)$$

Where nXn is the size of the subband under consideration.

Thus $$\hat{T}_B(\hat{\sigma}_x)=\frac{\hat{\sigma}^2}{\hat{\sigma}_X} \quad (2.14)$$

Where

$$\hat{\sigma}_x=\sqrt{\max\left(\hat{\sigma}_Y^2-\hat{\sigma}^2,0\right)} \quad (2.15)$$

In the case that $\hat{\sigma}^2 > \hat{\sigma}_Y^2, \hat{\sigma}_x$ is taken to be zero, i.e, $\hat{T}_B(\hat{\sigma}_x)$ is ∞, or, in practice, $\hat{T}_B(\hat{\sigma}_x)=\max\left(|Y_{ij}|\right)$ and all coefficients are set to zero.
To summarize, Bayes Shrink performs soft thresholding, sub band dependent threshold,

$$\hat{T}_B(\hat{\sigma}_x)=\frac{\hat{\sigma}^2}{\hat{\sigma}_X} \quad (2.16)$$

The reconstruction using Bayes Shrink is smoother and more visually appealing than the one obtained using

Sure Shrink. This not only validates the approximation of the wavelet Coefficients to the GGD but also justifies the approximation to the threshold to a value independent of β.

5.5.7 Image Reconstruction:

Image reconstruction can be carried out by the following procedure:

(i) Lets sample the data by a factor of two on all the four sub bands at the coarsest scale

(ii) Filter the sub bands in each dimension.

(iii) Sum the four filtered sub bands to reach the low-low sub band at the next finer scale.

This above process can be repeated until the image is fully reconstructed.

$$\int_0^{r_2} F(r,f)\,dr\,df = [\sigma r_2/(2m_0)]$$
$$\times \int_0^{\infty} \exp(-\lambda|z_j - z_i|)\lambda^{-1} J_1(\lambda r_2) J_0(\lambda r_i)\, d\lambda. \quad (2$$

6. Results

Figure 1: Original Image

Figure 2: Noisy Image

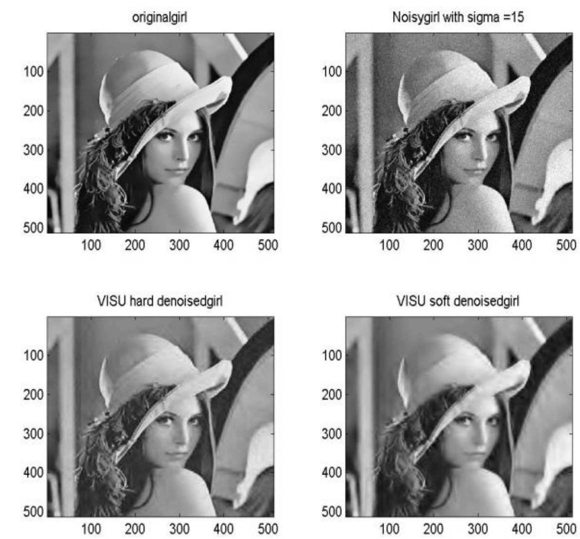

Figure 3: Visu Shrink

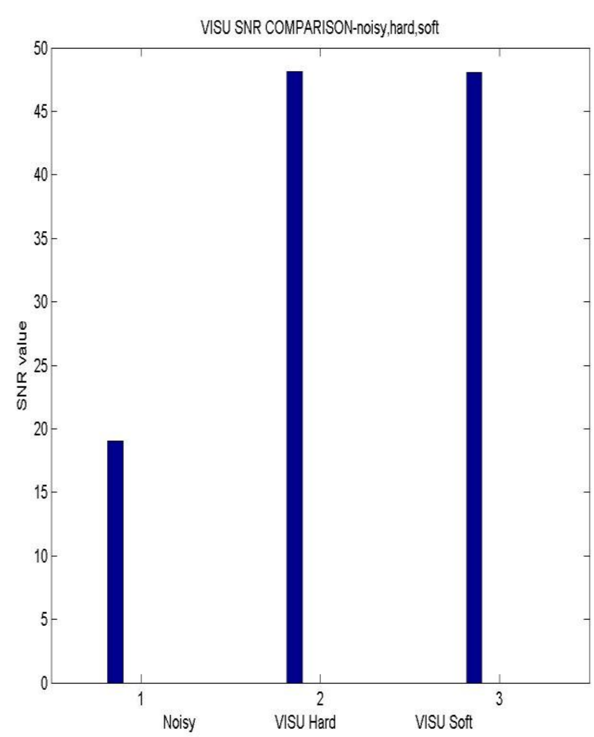

Figure 4: SNR of Visu shrink

Figure 5: Bayes Shrink

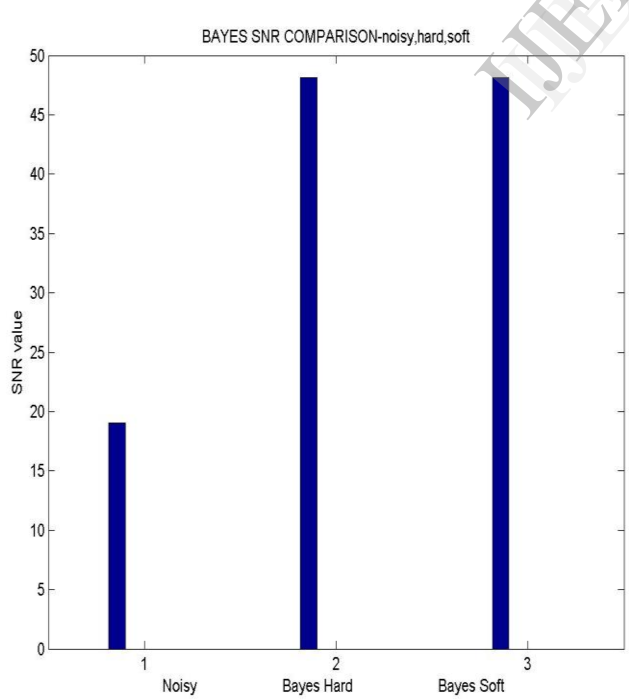

Figure 6: SNR of Bayes shrink

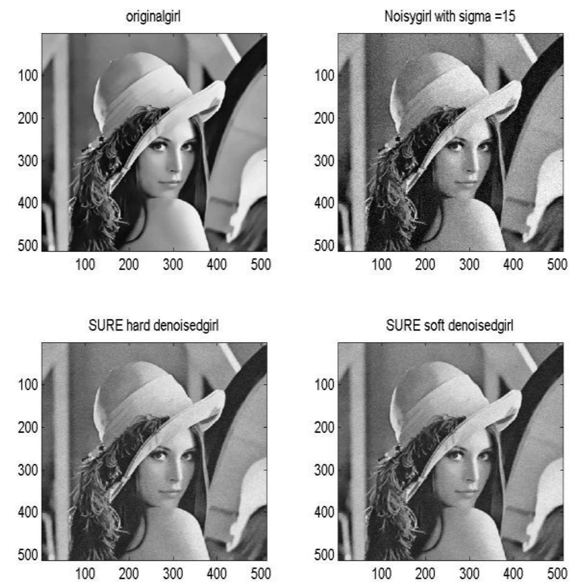

Figure 7: Sure Shrink

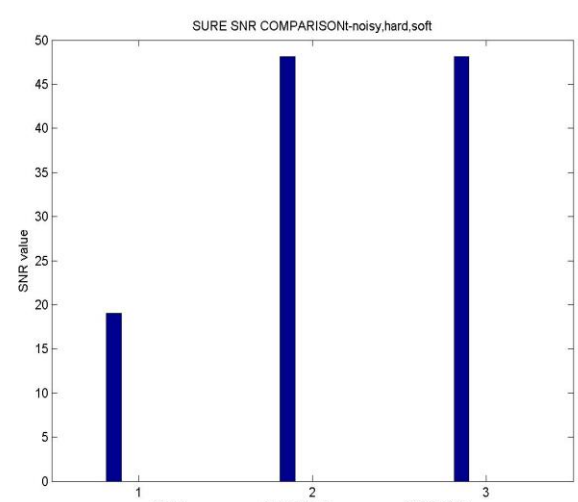

Figure 8: SNR of SURE shrink

7. Conclusion

We have seen that wavelet thresholding is an effective method of denoising noisy signals.Then we investigated many soft and hard thresholding schemes using Visu Shrink, Sure Shrink and Bayes Shrink for denoising images. It was found that sub band adaptive thresholding performs better than a universal thresholding. Among these, Bayes Shrink gave the best results. This validates the assumption that the Generalized Gaussian Distribution (GGD) is a very good model for the wavelet coefficient distribution in a sub band.

An important point to be noted is that although Sure Shrink performed worse than Bayes Shrink, it adapts well to sharp discontinuities in the signal. This was not evident in the natural images that were used for testing purpose specially while comparing the performance of these algorithms on artificial images with discontinuities (such as medical images).

Image denoising using thresholding by wavelets provides an extension for research into fast and robust multi-frame super resolution, exemplar-based inpainting and image feature extraction. Wavelet transform can also be used in the analysis and synthesis of multi-scale models of stochastic processes. The concept of wavelet transform finds an important application in speech coding, communications, radar, sonar, denoising, edge detection and feature detection

8. References

[1] Digital Image Processing, 3rd edition, by Rafael C.Gonzalez, Richard E.Woods ,Pearson Publications

[2] Digital Image Processing using MATLAB, 2nd edition, by Rafael C.Gonzalez,

[3] 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.

[4] David L Donoho. De-noising by soft thresholding. IEEE Transactions on Information Theory, 41(3):613–627, May 1995.

[5] Iain M.Johnstone David L Donoho. Adapting to smoothness via wavelet shrinkage. Journal of the Statistical Association, 90(432):1200–1224, Dec 1995.

[6] David L Donoho. Ideal spatial adaptation by wavelet shrinkage . Biometrika, 81(3):425–455,August 1994.

[7] Maarten Jansen. Noise Reduction by Wavelet Thresholding ,volume 161. Springer Verlag, United States of America, 1 edition, 2001.

[8] Carl Taswell. The what, how and why of wavelet shrinkage denoising. Computing in Science and Engineering , pages 12–19, May/June 2000.

[9] Sachin D Ruikar and Dharmpal D Doye . Wavelet based image denoising technique . (IJACSA) International Journal of Advanced Computer Science and Applications,Vol. 2, No.3, March 2011