Denoising of an Image Using Discrete Stationary Wavelet Transform and Various Thresholding Techniques

Abdullah Al Jumah
College of Computer Engineering & Science, Salman bin Abdul Aziz University, Al-Kharj, Saudi Arabia.
Email: syedamjadali.lords@gmail.com

Received August 13th, 2012; revised September 14th, 2012; accepted September 29th, 2012

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 denoising due to its properties such as multi-resolution. The problem of estimating an image that is corrupted by Additive White Gaussian Noise has been of interest for practical and theoretical reasons. Non-linear methods especially those based on wavelets have become popular due to its advantages over linear methods. Here I applied non-linear thresholding techniques in wavelet domain such as hard and soft thresholding, wavelet shrinkages such as Visu-shrink (non-adaptive) and SURE, Bayes and Normal Shrink (adaptive), using Discrete Stationary Wavelet Transform (DSWT) for different wavelets, at different levels, to denoise an image and determine the best one out of them. Performance of denoising algorithm is measured using quantitative performance measures such as Signal-to-Noise Ratio (SNR) and Mean Square Error (MSE) for various thresholding techniques.

Keywords: Wavelet; Discrete Wavelet Transform; Wavelet Packet Transform; Stationary Wavelet Transform; Thresholding; Visu Shrink; SURE Shrink; Normal Shrink; Mean Square Error; Peak Signal-to-Noise Ratio

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 keeping sharp transitions (i.e. edges) [1]. 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 de-noising. Wavelet denoising attempts to remove the noise present in the signal while preserving the signal characteristics, regardless of its frequency content.

Wavelet thresholding [2-5] (first proposed by Donoho) is a signal estimation technique that exploits the capabilities of wavelet transform for signal denoising. In our project, the wavelet thresholding techniques are applied to an image. 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. Figure 1 shows the block diagram of denoising using Wavelet transformation and thresholding techniques.

Denoising Procedure:

The procedure to denoise an image is given as follows:

De-noised image = W⁻¹ [T{W (Original Image + 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 thresholded image to get a denoised image in spatial domain.

2. Theoretical Aspects of Image Denoising Techniques

2.1. Discrete Wavelet Transform (DWT) [6-8]

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
Denoising of an Image Using Discrete Stationary Wavelet Transform and Various Thresholding Techniques

Figure 1. Block diagram of de noising using wavelet transformation and thresholding techniques.

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: [9,10]

\[ y[n] = (r * g)[n] = \sum_{k=-\infty}^{\infty} x[k] g[n-k] \]

\[ y_{low}[n] = \sum_{k=-\infty}^{\infty} x[k] g[2n-k] \]

\[ y_{high}[n] = \sum_{k=-\infty}^{\infty} x[k] h[2n-k] \]

\[ y \downarrow k \] is the down sampling operator. The Discrete Wavelet Transform provides sufficient information both for analysis and reconstruction of the original signal, with a reduction in the computation time.

Sub-Band Coding

Sub-band coding is a method for calculating the Discrete Wavelet Transform. The whole sub-band process consists of a filter bank, and filters of different cut-off 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. An example of this process is illustrated in Figure 2. 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.

Where, \( c_{A_x} \) is the approximation coefficients at decomposition level \( x \), \( c_{D_x} \) is the detail coefficients at decomposition level \( x \). S is the original signal. From Figure 2, you can see the original signal is broken down into different levels of decomposition. In the above case, it is a 3-level decomposition. Every time the newly scaled wavelet is applied to the signal, the information captured by the coefficients remains stored at that level. Thus the remaining information contains the higher frequencies of the signal, if the scaling factor decreases.

2.2. Stationary Wavelet Transform

The Stationary wavelet transform (SWT) is similar to the DWT except the signal is never sub-sampled and instead the filters are up sampled at each level of decomposition.

Each level’s filters are up-sampled versions of the previous as shown in Figure 3.
The SWT is an inherent redundant scheme, as each set of coefficients contains the same number of samples as the input. So for a decomposition of N levels, there is a redundancy of 2N.

Figure 4 shows the decomposition of Discrete and Stationary wavelet transform. The Discrete Wavelet Transform (DWT) \[11,12\] is the simplest way to implement MRA. It necessitates a decimation by a factor 2N, where N stands for the level of decomposition, of the transformed signal at each stage of the decomposition. As a result, DWT is not translation invariant which leads to block artifacts and aliasing during the fusion process between the wavelet coefficients. For this reason, we use the Stationary Wavelet Transform (SWT) (Holschneider, 1988). For the SWT scheme the output signals at each stage are redundant because there is no signal downsampling; insertion of zeros between taps of the filters are used instead of decimation. Figure 5 shows the decomposition of an image using SWT at level 1.

3. Thresholding Techniques

Thresholding \[13,14\] 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.

3.1. 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 denosing 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 thresholding.

\[
T^\text{Hard}_\tau(Y[m,n]) = \begin{cases} 
Y[m,n], & |Y[m,n]| > \tau \\
0, & |Y[m,n]| \leq \tau 
\end{cases}
\]

where, \( \tau \) is the threshold value.

From Figure 6, 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 7 shows Hard thresholding using dB5 wavelet.

### 3.2. Soft Thresholding

With a slight modification to the hard thresholding approach, a method known as soft thresholding can be created as shown in Figure 8.

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

where, \( T \) is the threshold value.

From Figure 9, the original image had more noise in
the bottom half compared to the top half. Since soft thresholding is a global operation, in the sense the entire image is used for the denoising process, it cannot concentrate on just the lower half of the image. But, if the denoised image has to be processed again, then the top half of the image would be over processed, and defects, such as blurring, can be introduced. Some original details of the image is removed along with the noise. This is because the noise obscured most of the small magnitude values that result from the original signal. Consequently, when thresholding is applied, it removes many of the transform values of the original signal, which are needed for accurate approximation. To overcome this problem wavelet shrinkages are used.

3.3. Shrinkages

3.3.1. VISU Shrink

- VisuShrink is thresholding by applying the universal threshold proposed by Donoho and Johnstone.
- This threshold is given by
  \[ \sigma \sqrt{2 \log M} \]
  where \( \sigma \) is the noise variance and \( M \) is the number of pixels in the image.
- For denoising images, VisuShrink is found to yield an overly smoothed estimate.

3.3.2. SURE Shrink

- SURE Shrink [15] 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 \( \{X_i; i = 1, \ldots, d\} \)
- For the soft threshold estimator \( \hat{X}_i = \eta(X_i) \)
  we have
  \[ SURE(t, X) = d - 2 \# \{i; |X_i| \leq t\} + \sum_{i=1}^{d} \min(\{X_i, t\})^2 \]
  - Select threshold \( t^* \) by
    \[ t^* = \arg \min SURE(t, X) \]

3.3.3. BAYES Shrink

- Bayes Shrink is an adaptive data-driven threshold for image de-noising via wavelet soft-thresholding.
- We assume generalized Gaussian distribution (GGD) for the wavelet coefficients in each detail sub-band.
- We then try to find the threshold \( T \) which minimizes the Bayesian Risk.

3.4. Experimental Results

<table>
<thead>
<tr>
<th>Wavelet</th>
<th>DB5 GH L1</th>
<th>HAAR GH L1</th>
<th>COIF5 GH L1</th>
<th>SYM4 GH L1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Threshold</td>
<td>B.T</td>
<td>A.T</td>
<td>B.T</td>
<td>A.T</td>
</tr>
<tr>
<td>MSE</td>
<td>0.13</td>
<td>2.14E-13</td>
<td>0.13</td>
<td>2.60E-16</td>
</tr>
<tr>
<td>SNR</td>
<td>16.37</td>
<td>252.36</td>
<td>15.95</td>
<td>310.88</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Wavelet</th>
<th>DB5 GS L1</th>
<th>HAAR GS L1</th>
<th>COIF5 GS L1</th>
<th>SYM4 GS L1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Threshold</td>
<td>B.T</td>
<td>A.T</td>
<td>B.T</td>
<td>A.T</td>
</tr>
<tr>
<td>MSE</td>
<td>0.13</td>
<td>2.37E-13</td>
<td>0.12</td>
<td>2.85E-16</td>
</tr>
<tr>
<td>SNR</td>
<td>16.37</td>
<td>252.49</td>
<td>16.86</td>
<td>310.88</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Wavelet</th>
<th>DB5 VS L1</th>
<th>HAAR VS L1</th>
<th>COIF5 VS L1</th>
<th>SYM4 VS L1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Threshold</td>
<td>B.T</td>
<td>A.T</td>
<td>B.T</td>
<td>A.T</td>
</tr>
<tr>
<td>MSE</td>
<td>0.12</td>
<td>2.36E-13</td>
<td>0.12</td>
<td>2.85E-16</td>
</tr>
<tr>
<td>SNR</td>
<td>17.05</td>
<td>252.54</td>
<td>15.95</td>
<td>310.88</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Wavelet</th>
<th>DB5 VH L1</th>
<th>HAAR VH L1</th>
<th>COIF5 VH L1</th>
<th>SYM4 VH L1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Threshold</td>
<td>B.T</td>
<td>A.T</td>
<td>B.T</td>
<td>A.T</td>
</tr>
</tbody>
</table>

Copyright © 2013 SciRes.
Denoising of an Image Using Discrete Stationary Wavelet Transform and Various Thresholding Techniques

<table>
<thead>
<tr>
<th>MSE</th>
<th>0.12</th>
<th>2.37E–13</th>
<th>0.13</th>
<th>2.83E–16</th>
<th>0.11</th>
<th>6.01E–10</th>
<th>0.12</th>
<th>4.80E–14</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNR</td>
<td>16.87</td>
<td>252.48</td>
<td>16.45</td>
<td>310.95</td>
<td>17.61</td>
<td>184.42</td>
<td>16.67</td>
<td>310.95</td>
</tr>
</tbody>
</table>

**BAYES SHRINK**

<table>
<thead>
<tr>
<th>Wavelet</th>
<th>DB5 BS L1</th>
<th>HAAR BS L1</th>
<th>COIF5 BS L1</th>
<th>SYM4 BS L1</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE</td>
<td>0.11</td>
<td>5.60E–13</td>
<td>0.11</td>
<td>2.86E–16</td>
</tr>
<tr>
<td>SNR</td>
<td>17.37</td>
<td>245.04</td>
<td>17.37</td>
<td>310.87</td>
</tr>
</tbody>
</table>

**SURE SHRINK HARD**

<table>
<thead>
<tr>
<th>Wavelet</th>
<th>DB5 SH L1</th>
<th>HAAR SH L1</th>
<th>COIF5 SH L1</th>
<th>SYM4 SH L1</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE</td>
<td>0.13</td>
<td>2.41E–13</td>
<td>0.12</td>
<td>2.84E–16</td>
</tr>
<tr>
<td>SNR</td>
<td>16.37</td>
<td>252.36</td>
<td>17.81</td>
<td>310.92</td>
</tr>
</tbody>
</table>

**SURE SHRINK SOFT**

<table>
<thead>
<tr>
<th>Wavelet</th>
<th>DB5 SS L1</th>
<th>HAAR SS L1</th>
<th>COIF5 SS L1</th>
<th>SYM4 SS L1</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE</td>
<td>0.12</td>
<td>2.36E–13</td>
<td>0.11</td>
<td>2.83E–16</td>
</tr>
<tr>
<td>SNR</td>
<td>17.32</td>
<td>252.53</td>
<td>81</td>
<td>310.95</td>
</tr>
</tbody>
</table>

**NORMAL SHRINK**

<table>
<thead>
<tr>
<th>Wavelet</th>
<th>DB5 NS L1</th>
<th>HAAR NS L1</th>
<th>COIF5 NS L1</th>
<th>SYM4 NS L1</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE</td>
<td>0.13</td>
<td>1.05E–12</td>
<td>0.12</td>
<td>8.44E–16</td>
</tr>
<tr>
<td>SNR</td>
<td>15.93</td>
<td>239.62</td>
<td>16.67</td>
<td>301.48</td>
</tr>
</tbody>
</table>


![Haar at level 1 SNR before and after thresholding](image)

Figure 10. Graph for Haar at level 1 for SNR before & after thresholding.
Comparison of SNR before and after thresholding for various wavelets at level 1 is shown in Figures 10-13. The Figure 14 shows image outputs for dB5 at level 1 for global soft thresholding.

3.5. Conclusion

DWT is translation varianopusion process between the wavelet co-efficient, whereas DSWT is Translation Invariant. In DSWT artifacts and aliasing are less than compared to DWT. This is the reason why DSWT is preferred over DWT. We have applied Additive White Gaussian Noise to an original image (kid), then DSWT is applied to get decomposed wavelet co-efficients to which various threshold techniques are applied for different wavelets. Inverse DSWT is applied to get reconstructed denoised image. In this paper, we have compared different wavelets such as Daubachies, Haar, Coiflet, Symlet with various threshold techniques such as default global hard and soft, VISU shrink soft and hard, Bayes shrink soft and hard, SURE shrink soft and hard and Normal shrink and measured the parameters such as Mean Square Error (MSE) and Signal-to-Noise Ratio (SNR). After comparison, it is found that MSE for HAAR global hard wavelet threshold is the least among all. SNR for HAAR SURE shrink soft level 1 is the maximum and the best among all.

3.6. Future Scope

Nonetheless, there is always room for improvement. Multiwavelets are relatively a new subject of study. Most
current filters available have two, three or fourth order of approximation. Future construction methods may add even higher order of approximation, while preserving the desirable features of current methods. It most likely result in multi-filters that perform even better in image denoising & compression applications. Moreover the multiwavelet systems available presently have the multiscaling and multiwavelet coefficients which are $2 \times 2$ matrices. There is a possibility that in future many more multiwavelet systems might be developed with matrix coefficients with higher order, which could provide even better results in the field of image denoising & compression.

REFERENCES


