THRESHOLDING AND TRANSFORM ORTHOGONALITY EFFECTS IN WAVELET SHRINKAGE AND DENOISING

Dayong Zhou, Victor DeBrunner and Joseph Havlicek

School of Electrical and Computer Engineering
The University of Oklahoma
202 West Boyd Street, Room 219
Norman, OK 73019
{dayong,vdebrunn,joebob}@ou.edu

ABSTRACT

We develop a novel algorithm based on the method of wavelet shrinkage for denoising images that combines the best features from three recently published methods [1]-[3]. In the process of developing the algorithm, we have noted several key aspects where improvements can be made and hold the potential for significant performance improvements. In exploring these important issues, we fully investigate the choices of the threshold used in defining the nonlinear filter used for denoising, as well as the impact of estimation errors on the algorithm performance. We also put some perspective on the impact of employing non-orthogonal representations. Our proposed algorithm is computationally inexpensive.

1. INTRODUCTION

Unlike the Fourier transform, the wavelet transform gives a multiresolution analysis of a signal. It is used widely in signal processing applications such as denoising and coding. The wavelet shrinkage (denoising) method introduced by Donoho and Johnstone [4] is a popular method for image denoising. In this approach, large transform coefficients are assumed to be associated with the signal while small transform coefficients are assumed to be associated with the noise. However, this approach exhibits spurious oscillations and other visual artifacts. Lang, *et al.*, proposed a denoising algorithm using an undecimated wavelet transform (UDWT) [1]. The shift invariance of the UDWT appears to improve the denoising performance [3]. The key factor in performance, as determined by all researchers to this point, is the thresholding used to do the filtering. With this in mind, Xu, *et al.*, proposed the additional use of a spatially-selective nonlinear filter [2]. Then, Pan, *et al.*, [3] modified the Xu algorithm slightly in its method of choosing the nonlinear filtering to yield a slightly improved performance. An alternative version (there are two methods in the paper) given by Pan is in most aspects identical to the method from the Lang paper, with more detail.

Our simulation and analysis show that both the Xu and Pan methods are sensitive to the noise power estimate used in defining the nonlinear filter used to do the denoising. In this paper, we introduce a new spatially-selective noise filter based on the UDWT (possibly non-orthogonal) that incorporates ideas from these three papers. We will see that because of this basis, our proposed method robustly improves the denoising effect. Its performance is in this aspect similar to the performance of the Lang method, with slight improvements that result from our threshold selection method based on the Xu and Pan methods. Additionally, our analysis has indicated that improvements in the non-orthogonal basis selection method could significantly improve the denoising performance even further. We start by examining some details.

2. PREVIOUS WORK – SOME DETAILS

Xu [2] proposed an algorithm called the spatially-selective noise filtration technique that uses the nonorthogonal wavelet introduced by Mallat and Zhong [5]. The method extracts the signal information at each scale through the direct correlation of the coefficients at several adjacent scales. We have the inter-scale correlation

$$C_2(m,n) = \prod_{i=0}^{2} W(m+i,n), \quad n = 1,2,\ldots,N, \tag{1}$$

where $W(m,n)$ are the wavelet transform detail coefficients of the signal x at level m and point n. The first step in the algorithm is to normalize $C_2(m,n)$

$$\hat{C}_2(m,n) = C_2(m,n)\sqrt{P_W(m)/P_C(m)}, \tag{2}$$

where

$$\begin{aligned} P_W(m) &= \sum_n W^2(m,n) \\ P_C(m) &= \sum_n C_2^2(m,n) \end{aligned} \tag{3}$$

The normalized inter-scale correlation $\hat{C}_2(m,n)$ is used as a dynamic intra-scale threshold function. Transform components larger than the threshold are called "signal" and removed. $\hat{C}_2(m,n)$ is then recomputed based on the remaining components. This renormalization of the inter-scale correlation lowers the threshold function, and more components are classified as signal. This iterated in-scale classification is stopped according to the estimated noise power. The devil in the details, of course, is determining the stopping point because accurate estimation of the noise power can be quite difficult. Stopping too early will leave significant signal out, while stopping too late allows too much noise. Simulation shows that this algorithm will generate more visually pleasing results when some signal is lost. To place this in some perspective, consider that in [4], the thresholding is done statically and directly in the transform domain, leading to inconsistent visual quality of the denoised images.

Pan observed in [3] that performance of the Xu algorithm can be improved by modifying the thresholds given in Eq. (2) according to

$$P_W(m) - \mathrm{th}(m)(N-K)\sigma_m^2 > 0.05 P_W(m) \qquad (4)$$

which stops the iteration slightly earlier. In (4), N is the total number of samples, K is the number of samples extracted as signal, and $1.2 \le \mathrm{th}(m) \le 1.8$ is a scale-dependent value. Rewriting, we have

$$N\frac{P_W(m)}{N-K} > \frac{\mathrm{th}(m)\sigma_m^2 N}{0.95} = (1.05\times \mathrm{th}(m))\sigma_m^2 N \qquad (5)$$

The left side is the estimated noise power based on the remaining wavelet coefficients. The right side is the estimated noise power based on noise variance, which is $P_W(m) > c\times$ estimation noise power at scale m, $c>1$.

Simulations indicate that this alteration in threshold generates better results when the noise power estimate is reliable. Simulation results based on the two algorithms are given in Figure 1 using a typical Lena image and added white Gaussian noise with MSE=600 out of 256 gray levels (about 10dB SNR). The horizontal axis gives the ratio of the estimated noise power to the true noise power; the vertical axis gives the percentage of noise extracted. We see that both algorithms can remove the noise effectively. However, the performance of both algorithms is very sensitive to noise power estimation errors.

Analysis shows this sensitivity is caused by the extraction process. If the noise power estimate is too low, as we mentioned above, then the algorithm will continue extracting noise after the signal information has been extracted. On the other hand, if the noise power estimate is too high, the algorithm will stop when significant signal information remains in the wavelet coefficients. This effect is most pronounced in the fine scales where the noise is dominant. These scales are critical for keeping sharp edges.

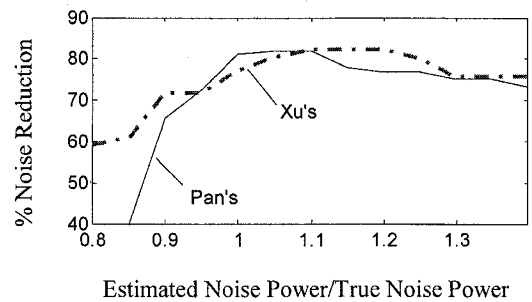

Figure 1. Noise reduction vs. noise power estimation

Since the Xu and Pan algorithms are sensitive to the noise power estimate, this paper considers methods to alleviate this concern. We introduce a spatially-selective filter based on the undecimated orthogonal wavelet transform (Details on efficient computation of the UDWT given in [6] are based on the original work pioneered in [7]) that is robust to the noise power estimation error.

The undecimated discrete wavelet transform generates an equal number of coefficients at all resolution levels, and so is overcomplete. The locality in the time-frequency plane provided by the wavelet allows us to denoise an image using the spatial information. Because the wavelet coefficients for each scale are computed as the output of FIR filter(s), computing the noise variance at each scale is straightforward. This transform is shift invariant and redundant. Traditional wavelet shrinkage using orthogonal, maximally-decimated wavelets sometimes exhibits visual artifacts such as Gibbs phenomena. Using the shift invariant wavelet in the shrinkage produces better results [6]. This is verified by the results given in [1] and [3], where the performance consisted of both visual and MSE metrics.

3. A NOISE ESTIMATION ERROR ROBUST SPATIAL SELECTIVE FILTER

Now that we have seen what others have done, we examine our tweaking of the thresholds in the nonlinear filtering.

3.1 Our Proposed Algorithm

We retain Eq. (1), in which $W(m,n)$ is the undecimated wavelet transformation detail coefficients of a noisy signal x at scale m and sample n. Some facets of the correlation remain unchanged. However, since we use a different wavelet decomposition than did Xu, we will distinguish between the correlations. Consequently, we introduce the notation $C_2'(m,n)$, to indicate that we are using the undecimated wavelet transform instead of the dyadic transform.

Our proposed algorithm is similar to the Pan spatially selective filter, but we use a different spatial selection criterion to extract the signal components. We give a brief description of our choice:

1. Take the undecimated wavelet transform of the signal.
2. Calculate $C_2'(m,n)$ at each scale.
3. If $C_2'(m,n)$ is larger than the estimated noise power, then select the component as a signal component.
4. Invert the undecimated wavelet transform off the signal components.

Like the Lang algorithm, our method does not require iteration. In the Xu and Pan algorithms, the threshold value is very critical to the performance of the algorithm. However, we are going to show in the following discussion that our algorithm is insensitive to the noise power estimate and therefore does not require a priori or precise knowledge of the noise process.

3.2. Noise power estimation insenstivity

$C_2'(m,n)$ is the product of wavelet transform coefficients at adjacent levels and may be modeled as a product of uncorrelated Gaussian processes ξ and γ. Thus ξ and γ are filtered sub-bands. Consequently, they are Gaussian, but not generally white. However, because the sub-bands are undecimated, they are orthogonal to each other. Of course, for Gaussian noise, orthogonal implies uncorrelated. As a result, we have the statistical properties:

$$\mathrm{E}\{C_2'(m,n)\} = E\{\xi\}E\{\gamma\} = 0 \tag{6}$$

$$Var\{C_2'(m,n)\} = Var\{\xi\}Var\{\gamma\} \tag{7}$$

$$f_\varsigma(C_2'(m,n)) = \int_{-\infty}^{\infty}\frac{1}{|\xi|}f_{\xi\gamma}\left(\xi,\frac{C_2'(m,n)}{\xi}\right)d\xi \tag{8}$$

It is difficult to calculate the relation in Eq. (8). However, Monte Carlo simulations show the PDF in figure 4. We find that the multiplication concentrates the distribution. As a result, if our threshold takes a relatively smaller value, most of contents of $C_2'(m,n)$ that are contributed from the noise can still be subtracted. On the other hand, edges are relatively unaffected because they appear at all scales. Here the locality of UDWT helps us immensely. $C_2'(m,n)$ is the multiplication of these two scales. So we can say that, through $C_2'(m,n)$, the noise information is suppressed while the signal information is amplified. Thus, we have that the difference of the parts of $C_2'(m,n)$ that are due to the noise and those due to the signal are much larger than the difference of the correlation to the wavelet coefficients themselves. Thus, our algorithm is robust to errors in the threshold value caused by mis-estimating the noise power.

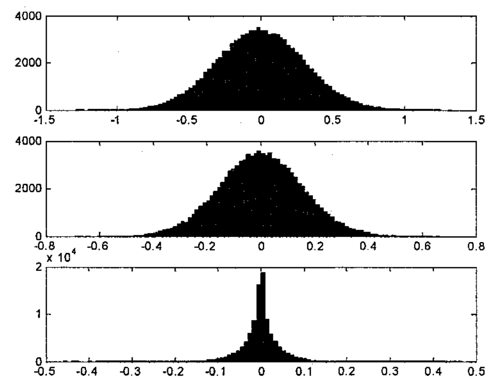

Figure 2. Sample distributions (level 5 top, level 6 middle, and product of the two at bottom)

3.3 Threshold selection

There are several ways to calculate the threshold. Here we provide two methods. In the first method we use the fact that the undecimated wavelet transform retains all spatial relationships. Consequently, to characterize the noise, we can estimate $C_2'(m,n)$ based on a region of the image where there is only noise present. In this case, we would have

$$threshold = c * \max(C_2'(m,n)),\ 0.5 \le c \le 0.8 \tag{9}$$

The second approach for calculating the threshold is a statistical method. In this case, we assume that the noise is AWGN, with distribution $\xi \sim N(0,\sigma^2)$. Then, we can use Eq. (6)-(8) to estimate the properties of $C_2'(m,n)$. In this case, we would have the threshold

$$threshold = c * \sigma_{\xi\gamma}^2,\ 0.5 \le c \le 0.8 \tag{10}$$

The thresholds in Eqs. (9) and (10) are constant in the correlation domain, but not in the wavelet transform domain. This is a significant difference from the Lang method [1].

4. SIMULATION

4.1 Sensitivity comparison

We simulate our algorithm and compare it with the other algorithms discussed in the introduction. We measure the reduced noise power with respect to the different noise power estimations. The simulation result is shown in Figure 3. We can see that our algorithm is very robust to the noise power estimation error. The "threshold" method shown is essentially the Lang method, or alternatively, the second variant given by Pan.

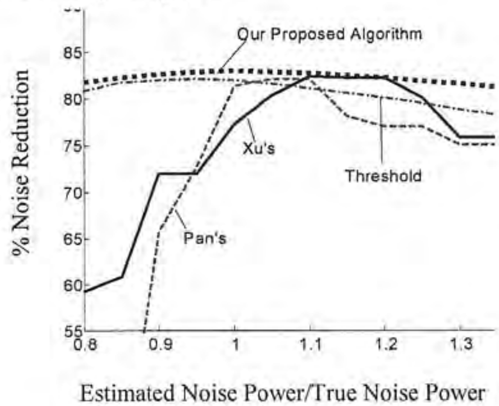

Figure 3. Sensitivity analysis with image Lena, σ^2=600

4.2 Visual comparisons

The above sensitivity simulation compares MSE performance which may not be indicative of visual quality. Accordingly, we compare the methods visually. Figure 4 shows the noisy image, the best result generated by the Xu algorithm (when the noise estimate is 1.2 times the true noise power), and our algorithm. In this case, the methods appear to be visually comparable. However, our result is significantly better than the typical results delivered by the Xu algorithm, as should be expected by examining Figure 3.

5. CONCLUSION

We have combined the best aspects of three previously described wavelet shrinkage algorithms used for image denoising. Our new approach is robust to misestimates of the noise power. As a result, we obtain MSE and visual performance that is equivalent to the best performances attainable by those previous algorithms and do so over a much wider range of power estimation errors. Moreover, we have noted that further improvements are possible by treating the non-orthogonal representation jointly in the image and wavelet domains, in a self-orthogonalizing matching pursuits context. This is our next task.

REFERENCES

[1] M. Lang, H. Guo, J. E. Odegaard, C. S. Burrus, and R. O. Wells, "Noise reduction using and undecimated discrete wavelet transform," *IEEE SP Lett.*, **3**, pp. 10-12, 1996.

[2] Y. Xu, J. B. Weaver, D. Healy, and J. Lu, "Wavelet transform domain filters: A spatially selective noise filtration technique," *IEEE T-IP*, vol. 3, no. 6, Nov 1994.

[3] Q. Pan, L. Zhang, and G. Dai "Two denoising methods by wavelet transform," IEEE T-SP, vol. 47, no. 12, Dec 1999.

[4] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation via wavelet shrinkage," *Biometrica*, **81**, pp. 425-455, 1994.

[5] S. Mallat and S. Zhong, "Characterization of signals from multiscale edges," *IEEE T-PAMI*, vol. 14, no. 7, Jul 1992.

[6] M. Jansen, "Noise reduction by wavelet thresholding" *Springer*, 2001.

[7] S. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE T-PAMI*, vol. 11, no. 7, pp. 674-693, 1989.

a) Picture before denoising

b) Denoised picture by our algorithm

c) Best denoised picture by Xu's algorithm

Figure 4. Visual comparisons