Denoising of imagery for inspection tasks using higher-order statistics

Samuel P. Kozaitis
Florida Institute of Technology
Department of Electrical and Computer Engineering
150 W. University Blvd.
Melbourne, FL 32901
kozaitis@fit.edu

ABSTRACT

We reduced noise in images using a higher-order, correlation-based method. In this approach, wavelet coefficients were classified as either mostly noise or mostly signal based on third-order statistics. Because the higher than second-order moments of the Gaussian probability function are zero, the third-order correlation coefficient may not have a statistical contribution from Gaussian noise. Using a detection algorithm derived from third-order statistics, we determined if a wavelet coefficient was noisy by looking at its third-order correlation coefficient. Using imagery of space shuttle tiles, our results showed that the minimum mean-squared error obtained using third-order statistics was often less than that using second-order statistics.

Keywords: correlation, denoising, Gaussian noise, higher-order statistics, wavelet transforms

1 INTRODUCTION

The wavelet transform has been used in denoising because it usually compacts a signal better than it does noise, and the noise and signal can often be separated due to this property. A transform coefficient can be seen as a cross-correlation between a basis function and an input signal. Therefore, a transform is a collection of cross-correlation coefficients between an input and the basis functions of the transform. To denoise a signal in the case of additive noise, noisy coefficients can be eliminated or the noise subtracted. Identifying coefficient values in the presence of noise is crucial to denoising.

Denoising methods are often based on estimates of the magnitude or power of the signal and noise present. Some popular soft thresholding methods in the wavelet domain have been developed for real signals. [1] Similar approaches have been developed for a variety of thresholding methods [2-4] including a hard threshold for nonorthogonal transforms [5], a method based on Bayes estimation [6], and a soft threshold for complex signals. [7] Some recent methods include those that address noise distributions that are not necessarily Gaussian by detecting outliers based on an iterative method [8] or robust regression [9]. In general, these methods detect values that strongly influence second-order statistics.

Promising results have been obtained with methods that resemble Wiener filtering. Because the Wiener filter is an optimal linear filter if the complete second-order statistics of the signal and noise are known, wavelet image coefficients are modeled as Gaussian random variables [10]. Using a spatially adaptive wavelet image model, wavelet coefficient variances are estimated before an approximate minimum mean squared error estimation procedure is used [11]. In such methods, it is important to accurately estimate the signal variance from the noisy

Wavelet Applications in Industrial Processing IV, edited by Frédéric Truchetet, Olivier Laligant,
Proceedings of SPIE Vol. 6383, 63830J, (2006) · 0277-786X/06/$15 · doi: 10.1117/12.686619

signal. Nearly arbitrary shaped windows have been used [12] as well as directional windows for different oriented subbands [13]. In all cases, when signal and noise powers are known, an estimate of signal is obtained using relatively low complexity.

More complicated methods often exploit multiple scale dependencies of the wavelet transform. By thresholding the products of wavelet coefficients between scales, edges may be extracted and noise reduced [14,15]. Using the evolution of wavelet coefficient magnitudes across scales and spatial clustering are other methods [16]. In another approach, the wavelet coefficients at the same spatial locations of two adjacent scales are represented as a vector, and the linear minimum mean square-error is applied to the vector [17]. The optimal wavelet is then determined from a library of wavelet bases.

In contrast to previous work, we determined whether wavelet coefficients were noisy based on higher-order correlations. Higher-order correlations are an extension of the more familiar second-order, cross-correlation function, but have the advantage of being insensitive to noise of unknown spectral density under certain conditions. [18] In our approach, we applied a third-order correlation detection method to identify wavelet coefficients that contained mostly signal [19,20]. Because the higher than second-order moments of the Gaussian probability function are zero, a third-order correlation coefficient will not have a statistical contribution from Gaussian noise. This paper gives the theory of detection of wavelet coefficients using higher-order correlations in the presence of additive noise and reports on its applications. After briefly discussing higher-order correlations, we present the main characteristics of higher-order denoising followed by a theoretical analysis when used in conjunction with wavelet transforms. Then, we apply this method to denoise images with Gaussian noise and compare the results to the minimum error obtained with second-order statistics.

2 HIGHER-ORDER CORRELATIONS

The correlation between two functions has often been used as a measure of their similarity. The conventional correlation function is second-order and is a special case higher-order correlation [18]. Although higher-order correlations have been used for many years, their use has been limited. The *n*th-order autocorrelation of the signal $f(x)$ is defined as

$$f_n(\tau_1, \tau_2, \ldots \tau_{n-1}) \equiv \sum_{k=0}^{N-1} f(x) f(\tau_1+x) f(\tau_2+x) \ldots f(\tau_{n-1}+x), \quad (1)$$

where the *n*th-order correlation is a function of $n - 1$ independent variables τ_n and N is the length of the signal. For $n = 2$, Eq. (1) becomes the second-order correlation of $f(x)$ which is the familiar autocorrelation function.

We primarily considered the third-order or triple correlation because it has the same advantages for our purpose and is easier to calculate than other higher-order correlations. The third-order correlation $n = 3$ of a one-dimensional function is a function of two variables. From Eq. (1) the third-order autocorrelation of $f(x)$ is

$$f_3(\tau_1, \tau_2) = \sum_{x=0}^{N-1} f(x) f(\tau_1+x) f(\tau_2+x), \quad (2)$$

where $f_3(\tau_1, \tau_2)$ is symmetric with respect to its variables τ_1 and τ_2.

The third-order correlation coefficient $f_3(0,0)$ can be found by sampling the triple correlation $f_3(\tau_1, \tau_2)$ at zero displacement where $\tau_1 = \tau_2 = 0$. From Eq. (2), the third-order correlation coefficient becomes

$$f_3(0,\ 0) = \sum_{x=0}^{N-1} f^3(x), \tag{3}$$

And shows that the third-order correlation coefficient $f_3(0, 0)$ of $f(x)$ can be calculated directly as the sum of the cubes of $f(x)$ from $x = 0$ - N - 1.

For a zero-mean symmetric distribution, the third-order correlation coefficient is zero. [18] Because the Gaussian distribution is symmetric about its mean, the third-order correlation of a zero-mean Gaussian process is zero.

3 HIGHER-ORDER SIGNAL DENOISING

This section describes our method for denoising using third-order correlations. Denoising typically involves calculating the wavelet transform of a signal, thresholding the wavelet coefficients, and calculating the inverse wavelet transform. Most methods calculate the threshold based on an analysis of the wavelet coefficients as indicated in Fig. 1, where $f(x)$ is the noise-free signal, $n(x)$ is the noise, $s(x)$ is the noisy signal, b_{jk} are the wavelet coefficients, j is the level of the transform, and k is the translation parameter. In many cases, the energy of the wavelet coefficients is used as a basis for thresholding or otherwise modifying their values, but there are other criteria that may be used [21]. After the inverse wavelet transform, the denoised signal $\hat{f}(x)$ should ideally be equal to the noise-free signal $f(x)$.

Our denoising approach was based on third-order correlations to identify wavelet coefficients comprised of mostly signal. We thought of wavelet coefficients as correlations between a noisy signal and wavelets at different scales and translations. The noise considered was zero-mean of unknown spectral density.

We first calculated the second-order cross-correlation *functions* between the noisy input signal and each scaled and translated wavelet and labeled the resulting functions as $b_{jk}(\tau)$. We then calculated the third-order autocorrelation coefficients from $b_{jk}(\tau)$. Using Eq. (3), the third-order autocorrelation coefficients of the second-order cross-correlations between wavelets and signal were described as

$$b_{3\,jk}(0,0) = \sum_{\tau=1}^{2m_j-1} \left(b_{jk}(\tau)\right)^3, \tag{4}$$

where m_j is the length of the wavelet at the *jth* scale, and the summation is performed only over the portion of the signal where the wavelet is supported. The block diagram of our approach is shown in Fig. 2. Because the wavelet coefficients b_{jk} represent single values in each of the functions $b_{jk}(\tau)$, the wavelet coefficients cannot be used to calculate the third-order correlation coefficients. To calculate the function $b_{jk}(\tau)$, the basis functions $w_{jk}(\tau)$ must be available. The third-order correlation coefficients $b_{3jk}(0,0)$ are then thresholded to select which wavelet coefficients are to be used in reconstructing the signal.

We considered the result of the second-order wavelet-signal cross-correlation as consisting of two parts. One part was the correlation between the input signal and the wavelet, and the second part was the correlation between the noise and the wavelet. The second-order correlation result was written as

$$b_{jk}(\tau) = fb_{jk}(\tau) + nb_{jk}(\tau), \tag{5}$$

where $fb_{jk}(\tau)$ and $nb_{jk}(\tau)$ represented the noise-free signal-wavelet correlation and the noise-wavelet correlation respectively. Substituting Eq. (5) into Eq. (4) and rearranging, the expression for the third-order autocorrelation coefficient can be written as [22],

$$b_{3\,jk}(0,0) = \sum_{\tau=1}^{2m_j-1} \left(fb_{jk}(\tau)\right)^3 + 3\sum_{\tau=1}^{2m_j-1} \left(fb_{jk}(\tau)\right)^2 nb_{jk}(\tau) + 3\sum_{\tau=1}^{2m_j-1} fb_{jk}(\tau)\left(nb_{jk}(\tau)\right)^2 + \sum_{\tau=1}^{2m_j-1} \left(nb_{jk}(\tau)\right)^3. \tag{6}$$

If we assume zero-mean noise, the second term in Eq. (6) will statistically approach zero as the length of the signal increases. If the noise has a symmetric distribution, then the last term will also approach zero. The third term is related to the product of the signal and the noise power. It can be minimized if the mean of the signal is set to zero over the region of correlation. The first term, the third-order correlation, is not due to noise and can be separated from the remaining terms using a threshold T as

$$\hat{b}_{jk} = \begin{cases} b_{jk} & if \quad b_{3\,jk}(0,0) \geq T \\ 0 & otherwise \end{cases}. \tag{7}$$

A third-order detection algorithm is used in this way to select wavelet coefficients that contain mostly signal.

4 ANALYSIS

The last term in Eq. (6) is the third-order correlation coefficient of the noise-wavelet correlation function. Each value of $nb_{jk}(\tau)$ can be seen as a sum of products, where the products are between the wavelet coefficients and samples of the noise $n(\tau)$. Considering uncorrelated identically distributed noise, the samples of $nb_{jk}(\tau)$ are random variables whose probability density functions (pdfs) can be written as

$$nb_{jk}(\tau) = \sum_{i=0}^{m_j-1} Xc(i), \tag{8}$$

where the $c_j(i)$'s are the coefficients of the wavelet at the jth scale and X is a random variable with a distribution of the noise signal. If X is zero-mean Gaussian distributed with variance σ_x^2, then the expression in Eq. (8) will be normally distributed with mean and variance $\mu_{nb} = 0$ and $\sigma_{nb}^2 = \sigma_x^2 \sum_{i=0}^{m_j - 1} c_j^2(i)$.

To complete the contribution of the noise-only term to the third-order correlation coefficient in Eq. (6), the samples in Eq. (8) must be cubed and summed. Cubing the samples in Eq. (8) produces another random variable. When a random variable is obtained from another random variable by a strictly increasing or decreasing deterministic function such as $Z = H(Y)$, then Z will have the density function [11] $p_Z(z) = p_Y(y)|dy/dz|$. With the pdf of Eq. (8) described as $p_Y(y)$ and $z = y^3$, the pdf of Z can be written as

$$p_n(z) = \frac{1}{3\sigma^3 z^{1/3}\sqrt{2\pi}} \exp - \left(\frac{z^{1/3}}{\sigma}\right)^2, \tag{9}$$

where $p_n(z)$ is not defined at $z = 0$.

The last term in Eq. (6) is a summation of $2m_j - 1$ random variables. Its pdf can therefore be written as the convolution of the densities as

$$p_{2m_i - 1}(z) = p_n(z) * p_n(z) * p_n(z) \ldots, \; 2m_i - 1 \text{ times.} \tag{10}$$

A plot of $p_{2m_i - 1}(z)$ is shown in Fig. 3 for values of $m_i = 2$, 4, 8, and 16. The results show that for small values of m_i, most values of the noise are shifted to zero. As m_i increases, the effect is reduced and the noise appears more Gaussian.

5 EXPERIMENT

We initially compared our method to the minimum mean-squared-error MSE using second-order statistics. We used three levels of the wavelet transform and, for second-order (conventional) denoising, found the minimum MSE possible for each SNR. The MSE was determined by varying a soft threshold value until the minimum MSE was found for each SNR. The same procedure was used for third-order denoising, but a hard threshold was used. A length 256 test signal is shown in Fig. 4(a) and the minimum MSE obtained with both second- and third-order processing is shown in Fig. 4(b) for $m_i = 6$ and Fig. 4(c) for $m_i = 16$. For all SNRs, the third-order result produced a lower MSE than the second-order case.

We used automatic thresholding of the third-order correlation coefficients by keeping constant the number of false rejections. This notion can be seen as a false discovery rate [24] (FDR) and can be written as

$$FDR = E\left\{\frac{N_1}{N_s\left(1 - \frac{1}{2^j}\right)}\right\}, \tag{11}$$

where N_S is the length of the input signal, N_1 is the number of false discoveries, and $E\{\ \}$ represents the expected value. The FDR can be determined by wavelet coefficients above a threshold. Using the monotonicity of the noise distribution $p_{2m_i - 1}(z)$, the FDR can be written as

$$FDR = 2[1 - \int_{-\infty}^{z_T} p_{2m_j - 1}(z)dz], \tag{12}$$

where $T = z_T$ is the threshold. The value of FDR as a function of z_T for different values of m_i was computed numerically and shown in Fig. 5. From this data, a threshold can be chosen to minimize the number of wavelet coefficients that have noisy third-order correlation coefficients.

We used an image of a damaged space shuttle tile as shown in Fig. 6 in our experiments. We estimated the noise statistics from the first level of the wavelet transform and choose a threshold value for a FDR = 0.1, and used Daubechies orthogonal minimum-phase wavelets in our experiments. In addition, we found that by denoising the rows and columns independently then averaging the results gave slightly improved results when compared to denoising rows followed by denoising columns. The results for m_i = 4,6,8, and 16 are shown in Fig. 7. The results show that the shortest wavelets gave the best performance. In addition, the results were compared to a conventional (universal) second-order denoising with m_i = 4. In all cases the third-order results gave better performance than the second order results.

6 CONCLUSION

A higher-order, correlation-based method was used for signal denoising. In this approach, a wavelet coefficient was classified as containing mostly noise or mostly signal based on higher-order statistics. This method should work best for noise distributions that are symmetric, because the third-order correlation coefficient of these distributions will be statistically zero. The method uses a hard threshold to eliminate noisy wavelet coefficients, but modifications to the approach to estimate the amount of noise in the third-order domain could possibly be used to improve results.

REFERENCES

[1] D. L. Donoho, and I. M. Johnstone, "Ideal spatial adaptation via wavelet shrinkage," *Biometrika*, **81**, 425-425 (1994).

[2] H-Y. Gao, "Wavelet shrinkage denoising using the nonnegative garrote," *J. Comput. Graph. Stat.*, **4**, 469-488 (1998).

[3] D. L. Donoho, "De-noising by soft thresholding," *IEEE Trans. on Inf. Theory*, **41**, 613-627 (1995).

[4] S. G. Chang, *et al.*, "Adaptive wavelet thresholding for image denoising and compression," *IEEE Trans. on Image Processing*, **9**(9), 1532-1546 (2000).

[5] Q. Pan, *et al.*, "Two denoising methods by wavelet transform," *IEEE Trans. on Signal Processing*, **47**(12), 3401-3406 (1999).

[6] H. A. Chipman, E. D. Kolaczyk, and R. E. McCulloch, "Adaptive Bayesian wavelet shrinkage, *J. Amer. Statist. Assoc*., **92**, 1413-1421 (1997).

[7] S. Sardy, "Minimax threshold for denoising complex signals with Waveshink," *IEEE Trans. on Signal Processing*, **48**(6), 1023-1028 (2000).

[8] R. Ranta, V. Louis-Dorr, C. Heinrich, and D. Wolf, "Iterative wavelet-based denoising methods and robust outlier detection," *IEEE Signal Processing Letters,* **12**(8), pp. 557-560 (2005).

[9] Z. Hou, and T. S. Koh, "Image denoising using robust regression," *IEEE Signal Processing Letters*, **11**(3), 243-246 (2004).

[10] S. LoPresto, K. Ramchandran, and M. T. Orchard, "Image coding based on mixture modeling of wavelet coefficients and a fast estimation-quantization framework," *in Proc. IEEE Data Compression Conf.*, Snowbird, UT, 221-230 (1997).

[11] M. K. Kivanc, I. Kozintsev, K. Ramchandran, and P. Moulin, "Low-complexity image denoising based on statistical modeling of wavelet coefficients," *IEEE Signal Processing Letters,* **6**(12), 300-303 (1999).

[12] I. K. Eom, and Y. S. Kim, "Wavelet-based denoising with nearly arbitrarily shaped windows," *IEEE Signal Processing Letters*, **11**(12), 937-940 (2004).

[13] P-L. Shui, "Image denoising algorithm via double local Wiener filtering with directional windows in wavelet domain," *IEEE Signal Processing Letters*, **12**(10), 681-684 (2005).

[14] Y. Lee, and S. P. Kozaitis, "Multiresolution gradient-based edge detection in noisy images using wavelet domain filters," *Optical Engineering*, **39**(9), 2405-2412 (2000).

[15] L. Zhang, and P. Bao, "Denoising by spatial correlation thresholding," *IEEE Trans. on Cir. & Sys. for Video Tech.,* **13**(6), 535-538 (2003).

[16] A. Pizurica, W. Phillips, I. Lemahieu, and M. Acheroy, "A joint inter- and intrascale statistical model for Bayesian wavelet based image denoising," *IEEE Transactions on Image Processing*, **11**(5), 545-557 (2002).

[17] L. Zhang, P. Bao, X. Wu, "Multiscale LMMSE-based image denoising with optimal wavelet selection," *IEEE Trans. on Cir. & Sys. for Video Tech.*, **15**(4), 469-481 (2005).

[18] J. M. Mendel, "Tutorial on higher-order statistics (spectra) in signal processing and system theory: theoretical results and some applications," *Proc. IEEE* **79**(3), 278-305, (1991).

[19] S. P. Kozaitis, and S. Kim, "Denoising using higher-order statistics," in *Independent Component Analysis, Wavelets, and Neural Networks*, T. Bell, M. V. Wickerhauser, H. H. Szu, Eds., Proc. SPIE 5102, paper 6 (2003).

[20] S. P. Kozaitis, R. H. Cofer, A. Ingun, "Reduction of multiplicative noise using higher-order statistics," in *Algorithms for Synthetic Aperture Radar Imagery X*, E. G. Zelnio, and F. D. Garber, Eds., Proc. SPIE 5095, paper 15 (2003).

[21] M. Jansen, Noise Reduction by Wavelet Thresholding, *Lecture Notes in Statistics* 161, Springer-Verlag: New York (2001).

[22] G. B. Giannakis, and M. K. Tsatsanis, *IEEE Trans. on Acoustics, Speech, and Signal Processing*, **38**, 1284-1296 (1990).

[23] K. S. Lii, and M. Rosenblatt, "Deconvolution and estimation of transfer function phase coefficients for non-Gaussian linear processes," *Ann. Stat.*, 1195-1208 (1982).

[24] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: A practical and powerful approach to multiple testing," *Journal of the Royal Statistical Society, Series B*, **57**, 289-300 (1995).

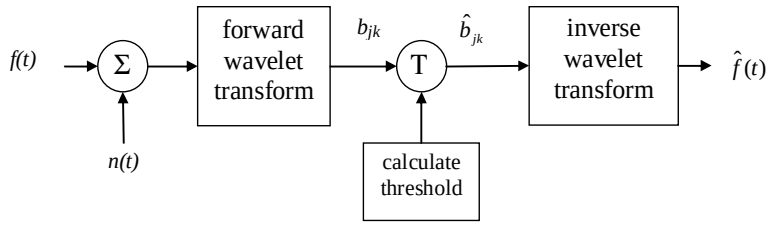


Figure 1 Conventional denoising approach.

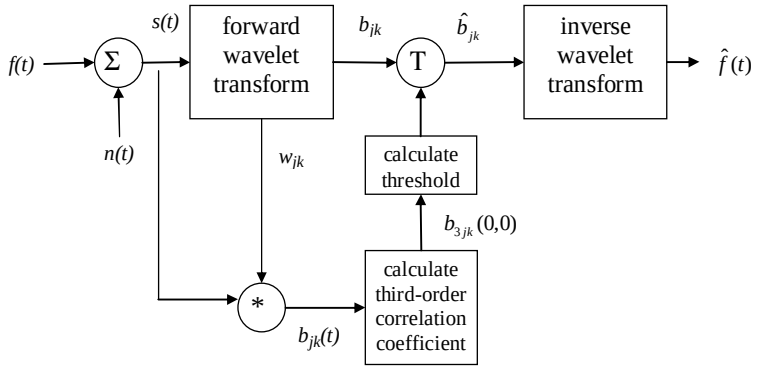


Figure 2 Third-order denoising approach.

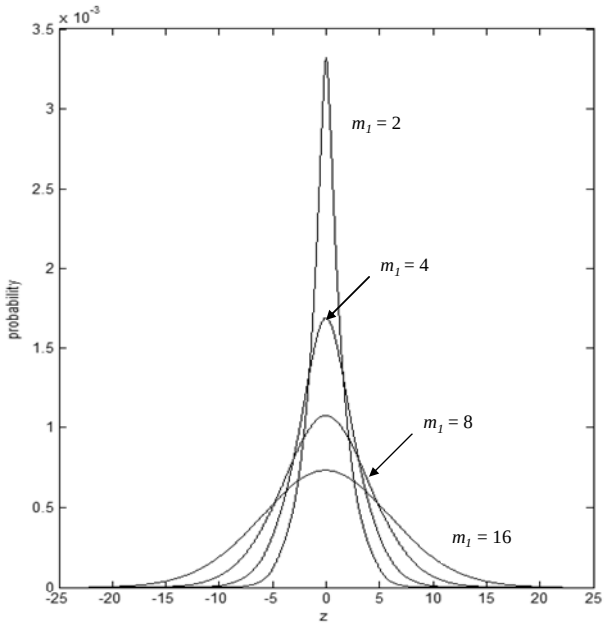
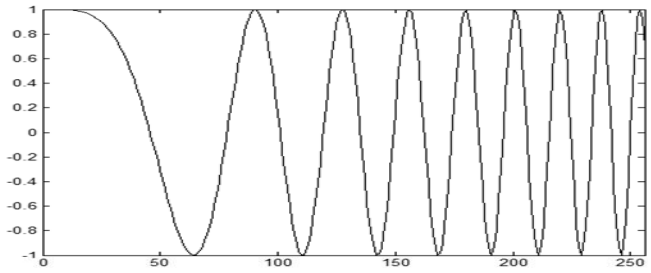
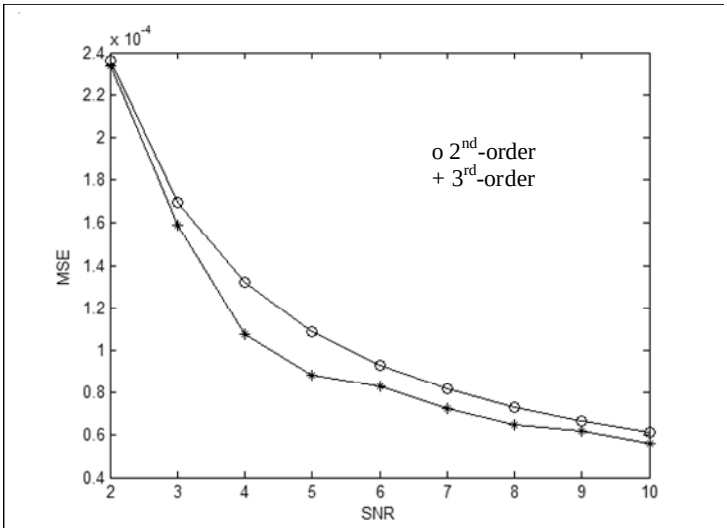


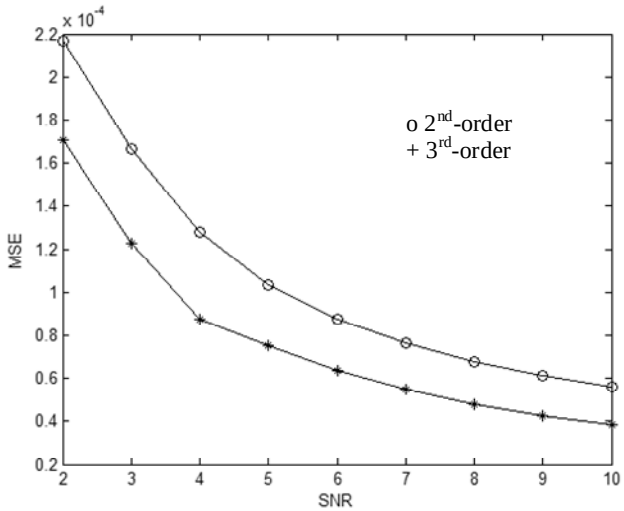
Figure 3 Probability density functions of $p_{2m_j - 1}(z)$ for wavelets of length 2, 4, 8, and 16.



(a)



(b)



(c)

Figure 4 Comparison of minimum MSE as a function of SNR using second- and third order denoisng with 3 levels for a length 256 signal (a) sample signal (b) results using $m_1 = 6$ (c) results using $m_1 = 16$.

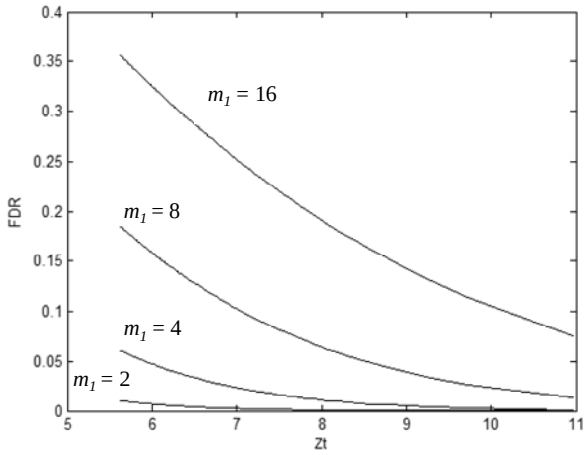


Figure 5 False discovery rate as a function of threshold for different length wavelets.

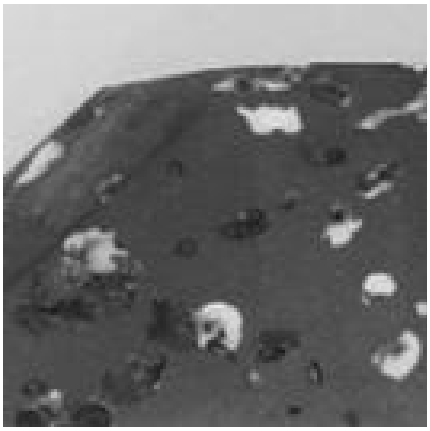


Figure 6 Image from space shuttle time used in experiments.

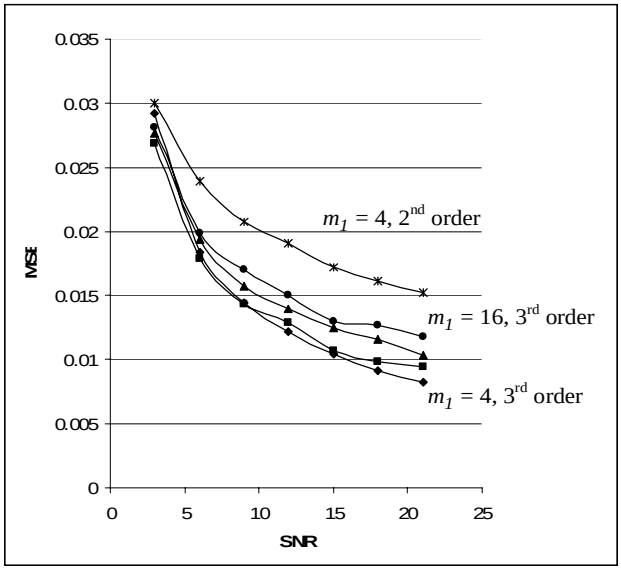


Figure 7 Comparison of MSE of 3rd-order and 2nd-order algorithms using image in Fig. 6.