Iranian Journal of Science & Technology, Transaction B, Engineering, Vol. 30, No. B6, pp 701-710
Printed in The Islamic Republic of Iran, 2006

DIGITAL IMAGE FILTERING IN WAVELET DOMAIN USING GENETIC PROGRAMMING*

M., EFTEKHARI AND H. J. EGHBALI**

Dept. of Computer Science and Engineering, Shiraz University, Shiraz, I. R. of Iran
Email: eghbali@shirazu.ac.ir

Abstract– Genetic Programming (GP) is a powerful machine learning technique derived from genetic algorithms. We used GP to generate a mathematical function for image denoising based on statistical features derived from detail sub-bands of wavelet transform (WT). The function obtained from GP for image denoising is not dependent to any parameters as represented in other image denoising methods based on WT. Results of the proposed image denoising method is compared to the VisuShrink soft threshold image denoising method, both perceptually and in terms of Peak Signal to Noise Ratio (PSNR).

Keywords– Genetic programming, wavelet transform, denoising, features, expressions, fitness, PSNR

1. INTRODUCTION

Wavelet Transform (WT) has become a popular tool for various image processing problems [1-5]. Signals can be represented as the frequency contents of local regions over a range of scales in wavelet domain. This is the most important feature of WT for analyzing the signals in one or more dimensions. Signal features exist in wavelet transform coefficients that make the signal analysis and synthesis much easier.

In this paper an image denoising method based on features estimated from the detail sub-bands of WT is considered. Let the image be defined by $f(i, j)$, $i, j = 0,1,..., N-1$ where, N is an integer power of 2. If $f(\cdot)$ is corrupted with additive white Gaussian noise, the noisy image observation $g(\cdot)$ will be given by:

$$g(i, j) = f(i, j) + n(i, j) \;.\; i, j = 1,2,...N \tag{1}$$

Where $n(\cdot)$ is white Gaussian noise with zero mean zero and variance σ_n^2, and is independent and identically distributed (i.i.d). The goal is to remove the noise from $g(i, j)$ and estimate $\hat{f}(i, j)$ which minimizes the mean square error (MSE) as given by:

$$MSE = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left[\hat{f}(i, j) - f(i, j)\right]^2 \tag{2}$$

In recent years there has been a fair amount of research on the image denoising based on WT [6-8]. The most popular methods are soft and hard thresholding introduced by Donoho and Johnson [9-11]. In wavelet domain, small wavelet coefficients more likely represent the noise, while large coefficients are a major feature of the original image. To decide which coefficient is small, a threshold is needed. Estimation of threshold is a major problem in this field. The widely used thresholding methods are VisuShrink [12], and SureShrink [13]. These two methods are based on minimizing the risk of Stein's risk estimation [14, 15]. The image denoising methods based on thresholding are discussed by Jansen Malfait and Bultheel

*Received by the editors November 22, 2004; final revised form May 6, 2006.
**Corresponding author

[16]. Wavelet shrinkage denoising using cross validation had been considered in [17-19]. Multiple hypotheses testing is suggested in [20-23] as another thresholding technique. These works face serious problems because of the large numbers of hypotheses needed to be tested simultaneously. Recently, block thresholding is utilized rather than term by term thresholding [24-29]. In block thresholding more information about neighbor coefficients are available. This leads to a better estimation of threshold. Blocks may be overlapping [30] or non-overlapping [31].

Soft thresholding employs a continuous function, while hard thresholding is a discontinuous function which causes some artifacts in denoised processing. Therefore, the soft thresholding method is preferred to hard thresholding [12, 32]. As mentioned above, the major problem for utilizing the soft thresholding method is the estimation of a good threshold. So in this work, the authors' intention is to propose an expression instead of the soft-thresholding function which is threshold independent. Genetic programming is commonly used to discover an optimum expression (program) for problems which need such expressions as their solutions [33-35]. Therefore, by using GP, a novel method of image denoising based on features drawn from the detail sub-band of WT is presented in this contribution. Our approach depends only on the statistical information of detail WT sub-bands that are easy to calculate.

This paper is arranged as follows. Section 2 discusses the Discrete Wavelet Transform (DWT). Wavelet shrinkage denoising is discussed in Section 3. Section 4 introduces the GP as a popular machine learning method to derive an expression for image noise removal. Section 5 describes simulation results. Finally, Section 6 draws the conclusions of this contribution.

2. THE DISCRETE WAVELET TRANSFORM

Wavelets are families of functions generated from a mother wavelet $\psi(t)$ by dilation and translation operations: $\psi_{m,n}(t) = 2^{m/2}\psi(2^m t - n)$ [36]. The mother wavelet is constructed from the scaling function $\phi(t)$, satisfying the following equation:

$$\phi(t) = \sqrt{2}\sum_{k=-\infty}^{\infty} h(k)\phi(2t-k) \tag{3}$$

The mother wavelet $\psi(t)$ is defined as:

$$\psi(t) = \sqrt{2}\sum_{k=-\infty}^{\infty} g(k)\phi(2t-k) \tag{4}$$

where *h(k)* and *g(k)* are a pair of discrete quadrature mirror filters that are related to each other as $g(k) = (-1)^k h(1-k)$. Here *h(k)* represents the low pass filter and *g(k)* the corresponding high pass filter.

The wavelet transform represents the decomposition of a function into a family of wavelet functions $\psi_{m,n}(t)$. In other words, using the wavelet transform, any arbitrary function f can be written as a superposition of wavelets [37]. Closely related with the wavelet transform is the multi-resolution analysis concept which is particularly appropriate for image analysis. So the one dimensional WT can easily be extended into two dimension.

The wavelet transform breaks an image down into four decimated images. They are subsampled by keeping every other pixel. The results consist of one image that has been high-pass (HP) filtered in both horizontal and vertical directions, one that has been high-pass filtered in the vertical and low-pass (LP) filtered in the horizontal, one that has been low-passed in the vertical and high-passed in the horizontal, and one that has been low –pass filtered in both horizontal and vertical directions.

Figure 1 shows the result of four bands (WT) on the Lena image based on Haar basis vectors [1]. The location of frequency bands in a four-band wavelet transform image is shown in Fig. 2.

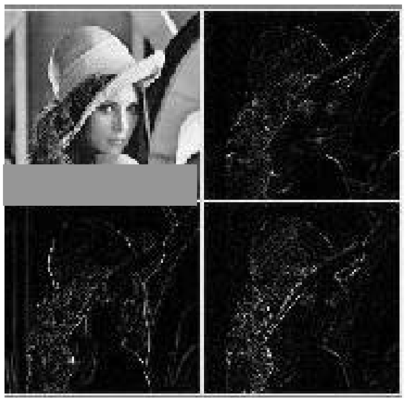

Fig. 1. Four bands WT of Lena image

LP/LP	**HP/LP**
LP/HP	**HP/HP**

Fig. 2. Locations of WT sub-bands

3. DENOISING BY WAVELET SHRINKAGE THRESHOLDING

Let $w(\cdot)$ and $w^{-1}(\cdot)$ denote the forward and inverse wavelet transform operators. Let $D(\cdot,\lambda)$ denote the denoising operator with soft threshold λ. The goal is to use wavelet shrinkage denoising on $g(\cdot)$ in order to recover $\hat{f}(\cdot)$ as an estimate of $f(\cdot)$ with minimum mean square error. Then, the three steps below summarize the procedure

$$y = w(g) \tag{5}$$

$$z = D(y,\lambda) \tag{6}$$

$$\hat{f} = w^{-1}(z) \tag{7}$$

where y is the detail sub-bands of DWT. z denotes the modified coefficients of DWT. The rule $D(y,\lambda) \equiv \mathrm{sgn}(y)\max(0, |y| - \lambda)$ defined by Donoho and Johnson is nonlinear soft thresholding. The operator D nulls all values of y *for* which $|y| \le \lambda$. Also all values of y *for* which $|y| > \lambda$ shrink toward the origin by an amount λ. It is the latter aspect that has led to D being called the shrinkage operator in addition to the soft thresholding operator.

For example, VisuShrink is a practical wavelet domain global threshold procedure. In this method the value of threshold is obtained from $\sigma_n\sqrt{2\log L}$. Here σ_n^2 is noise variance and L is the length of the data. In this paper a global denoising expression obtained by GP is introduced. So, the results of the proposed method are compared with the VisuShrink denoising method.

4. GP AS A TOOL TO GENERATE DENOISING EXPRESSIONS

Darwinian natural selection that involves both reproduction and the principle of the survival of the fittest causes biological species to robustly adapt to their environments. John Holland of the University of Michigan [38] introduced the algorithmic computer simulation of biological evolution that is called Genetic Algorithm (GA).

GAs are used for solving problems as a robust search method. Optimization problems are problems that can be solved by this nature-inspired type of algorithm.

Holland's works apparently illustrate the importance of using simulated genetic operators (crossover, mutation) [39]. He also presents, in his works, some methods that show how one can use genetic operators to give better performance for solving problems in adaptive systems.

GP was introduced by John Koza [40, 41] as a powerful machine learning method to solve problems requiring the discovery of a computer program or symbolic expressions as their solutions. GP follows the paradigms that are used in GAs.

Machine learning of a function, planning in artificial intelligence and robotics symbolic function identifications, time optimal control and a various range of problems [42, 43] require an optimal expression or computer program as their solutions.

The GP can be divided into a number of sequential steps defined below [41]:

1) Create a random population of symbolic expressions (programs) using symbolic functions in conjunction with both symbolic and constant terminals.
2) Evaluate each generated expression with a fitness value according to a predefined fitness function which measures the ability of the expression to solve the problem.
3) Using some predefined selection techniques to select some parents for recombination.
4) Genetically recombine the selected parents with the crossover operator to generate the new population.
5) Apply the mutation operator on this pool of new population.
6) Repeat Step 2 until predefined termination criteria are satisfied or a fixed number of generations are completed.
7) The solution to the problem is the expression with the best fitness within all the generations.

Generally in GP we need two sets to produce the expressions namely, function and terminal sets. In this work the set {Esqrt, Edivide, +,-,*} is defined as the function set and statistical features $\{\mu_1, \mu_2, \mu_3, \sigma_1, \sigma_2, \sigma_3\}$ are taken as the terminal sets, where μ_i and σ_i are the means and standard deviations of detail WT sub-bands respectively. An additional variable y (see Eq. (5)) which represents the detail sub-bands of WT is also used for the evaluation of the GP expression.

The terminals are the arguments of the functions. The combination of functions and terminals produce an expression that is a symbolic lisp like expression.

To find the optimum expression for image denoising, five noisy images of Lena with zero means and five different standard deviations were used. The values of five different standard deviations were, (σn_1=12.75, σn_2=20.4, σn_3=25.5, σn_4=38.25, σn_5= 63.75). Equation (8) is used as PSNR and the result of this equation was used to calculate the final fitness equation defined in (9).

$$PSNR = 20 \times \log_{10} \frac{\max_{i,j} \left| f\left(i,j\right) \right|}{\sqrt{MSE}} \tag{8}$$

$$\text{Fitness} = \sum_{k=1}^{5} \left| PSNR_k - \eta \right| \tag{9}$$

Where η is a constant that indicates an upper bound of the PSNRs. In this paper $\eta = 35$ is chosen, since, based on most researchers simulation results, the value of 35 is a right and proper estimation for PSNR. Equations (8) and (9) were used to evaluate the resulting optimum expression found by GP.

5. SIMULATION RESULTS

Three different test images were used. White Gaussian noise at five different levels of variances were generated in MATLAB using $\sigma \times randn$ and added to images. The aim is to obtain an expression for image denoising which results in the highest possible PSNR. Parameters used for the evaluation of the denoising expression in the wavelet domain are given in Table 1. In a mutation operation, a new individual is created by substituting a random subtree of the parent by a new randomly created tree [42]. Two-point cross-over is used in which two new individuals are created by swapping sub-trees of the two parents at 2 random points [42]. The selection method, called tournament selection, returns some random individuals chosen from the population using the tournament method, and duplicate individuals are allowed [42].

Table 1. The parameters of GP algorithm

Generation numbers	Population size	Selection method	Cross-over type	Cross-over probability	Mutation probability
300	100	Tournament	2-point	0.7	0.1

The best evaluated expression by GP is indicated in Eq. (10).

$$D(y,\sigma_3) = \text{Edivide}(\,y,\ \text{Esqrt}\ (\sigma_3)) \tag{10}$$

Where σ_3 is the variance of HP/HP WT sub-band, y represents DWT detail sub-bands and Esqrt and Edivide are defined below:

$$Esqrt\,(\theta) = \begin{cases} \sqrt{\theta} & if \quad \theta > 0 \\ 0 & otherwise \end{cases} \tag{11}$$

$$Edivide\,(\theta_1,\theta_2) = \begin{cases} \dfrac{\theta_1}{\theta_2} & if \quad \theta_2 \neq 0 \\ \infty & otherwise \end{cases} \tag{12}$$

Where θ is a scalar variable and θ_1, θ_2 are the arguments of *Edivide*. θ_1 is a matrix and is divided element-wise by the scalar θ_2. Also in (12) a large numerical value is considered to treat as the infinity. The overall results for different images are shown in Tables 2-7. Figure 3 shows a Lena image corrupted with a white Gaussian noise with Standard Deviation (SD) of 38.25. In all tables, the PSNRs of the proposed method are compared to the VisuShrink method in the wavelet domain.

Figures 4 and 5 indicate other test images to show the effectiveness of the proposed algorithm. Please note that different expressions can be obtained from GP. For example, a new expression found by GP is given in (13).

$$D(y,\sigma_1,\sigma_3) = \text{Edivide}(\,y,\ \text{Esqrt}\ (\sigma_1 \times \sigma_3)) \tag{13}$$

Tables 5-7 show the comparison of the proposed method with the VisuShrink denoising method for different test images using Eq. (13).

Fig. 3. a) Original image of Lena, b) Image corrupted by a gaussian noise with $\sigma n = 38.25$, c) VisuShrink method, d) proposed method

Table 2. PSNR values of the proposed method and VisuShrink
at 5 different SD of noise using Eq (10)

Method	σn				
	12.75	20.4	25.5	38.25	63.75
VisuShrink	25.8636	23.8774	22.7027	20.1952	16.7363
Proposed method	**26.5025**	**24.2771**	**22.9563**	**20.2697**	**16.7186**
Noisy image	24.3682	20.3758	18.4993	15.1606	11.2815

Table 3. PSNR values of the proposed method and VisuShrink
at 5 different SD of noise using Eq. (10)

Method	σn				
	12.75	20.4	25.5	38.25	63.75
VisuShrink	25.1284	23.4797	22.5778	20.5782	17.5810
Proposed method	**25.5841**	**23.9615**	**22.9311**	**20.7114**	**17.5920**
Noisy image	25.5366	21.4903	19.5950	16.2371	12.3753

Table 4. PSNR values of the proposed method and VisuShrink
at 5 different SD of noise using Eq. (10)

Method	σn				
	12.75	20.4	25.5	38.25	63.75
VisuShrink	22.8182	21.8532	21.1803	19.5050	16.8203
Proposed method	**23.8189**	**22.5912**	**21.7364**	**19.7765**	**16.8973**
Noisy image	25.1699	21.1016	19.1834	15.7830	11.9002

Table 5. **Lena image:** PSNR values of the proposed method and VisuShrink
at 5 different SD of noise using Eq. (13)

Method	σn				
	12.75	20.4	25.5	38.25	63.75
VisuShrink	25.8636	23.8774	22.7027	20.1952	16.7363
Proposed method	**25.9697**	**24.0355**	**22.8150**	**20.2298**	**16.7170**
Noisy image	24.3682	20.3758	18.4993	15.1606	11.2815

Table 6. **Barbara image:** PSNR values of the proposed method and VisuShrink
at 5 different SD of noise using Eq. (13)

Method	σn				
	12.75	20.4	25.5	38.25	63.75
VisuShrink	25.1284	23.4797	22.5778	20.5782	17.5810
Proposed method	**24.7226**	**23.4633**	**22.5938**	**20.5759**	**17.5612**
Noisy image	25.5366	21.4903	19.5950	16.2371	12.3753

Table 7. **Mandrill image:** PSNR values of the proposed method and VisuShrink
at 5 different SD of noise using Eq. (13)

Method	σn				
	12.75	20.4	25.5	38.25	63.75
VisuShrink	22.8182	21.8532	21.1803	19.5050	16.8203
Proposed method	**22.9227**	**22.0085**	**21.3160**	**19.5853**	**16.8445**
Noisy image	25.1699	21.1016	19.1834	15.7830	11.9002

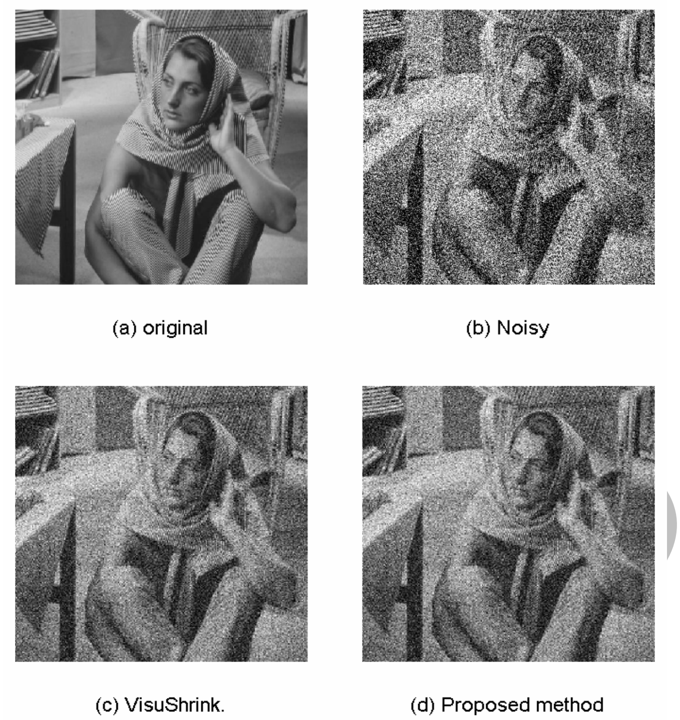

Fig. 4. a) Original image of Barbara, b) Image corrupted by a gaussian noise with σn = 63.75, c) VisuShrink, d) Proposed method

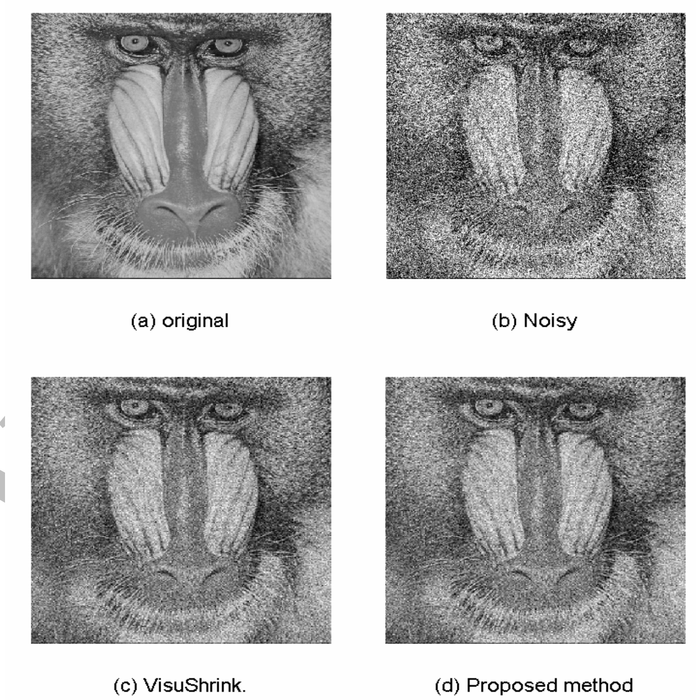

Fig. 5. a) Original image of Mandrill, b) Image corrupted by a gaussian noise with σn = 63.75, c) VisuShrink, d) proposed method

6. CONCLUSION

We have investigated and presented the application of GP for image denoising in the wavelet domain. GP produces and evolves symbolic equations for noise removal. The evolution process is robustly guided by a fitness function derived from the PSNR values.

As is apparent in statistical pattern recognition methods and also other applied statistical analysis approaches in engineering, statistical features are easy to compute and are available. So, by the use of such features, the problem of coming up with suitable and efficient features becomes easy. In this work the standard deviations and the means in detailed sub-bands of WT are used to derive expressions from GP for noise removal.

Although we have used only the Lena image to extract a suitable expression by GP, the resulting expression showed good behavior when tested on other images.

The proposed method indicates improved PSNRs as compared to the well-known VisuShrink denoising method in the wavelet domain.

REFERENCES

1. Mallat, S. (1999). *A wavelet tour for signal processing. 2nd Ed*. Newyork: Academic press.
2. Resnikoff, H. & Wells Jr., R. O. (2002). *Wavelet Analysis*. Spronger-Verlag, Newyork.
3. Zhang, J. K., Davidson, T. N. & Wong, K. M. (2003). Efficient design of orthogonal wavelet bases for signal representation. *IEEE Trans. Signal Proc*. *52*(7), 1983-1996.
4. Drori, I. & Lischinski, D. (2003). Fast multi-resolution image operations in the wavelet domain. *IEEE Trans. On Visualization and Computer Graphics*, *9*(3), 395-411.
5. Femandes, F., Selenick, I., Van Spaendonek, R. & Burrus, C. (2003). Complex wavelet transforms with all pass filters. *Signal Processing*, *83*(8), 1689-1706.
6. Eghbali, H. J. (2003). Adaptive digital image filtering in wavelet domain, *Machine Graphics & Vision*. *12*(2), 279-290.
7. Portilla, J., Strela, V., Wainwright, M. & Simoncelli, E. P. (2003). Image denoising using scale mixtures of gaussians in the wavelet domain. *IEEE Trans. on Image Processing*, *12*(11), 1338-1351.
8. Gilboa, G., Sochen, N. & Zeevi, Y. Y. (2004). Image enhancement and denoising by complex diffusion processes. *IEEE Trans. on Pattern Analysis and Machine Intelligence.26*(8), 1020-1036.
9. Donoho, D. L. (1995). Denoising by soft-thresholding, *IEEE Trans. Inf. Theory, 41*. 613-627.
10. Portilla, J. & Simoncelli, E. P. (2000). Image denoising via adjustment of wavelet coefficient magnitude correlation. *Seventh IEEE Int. Cong. On Image Proc., ICIP,* Vancouver, Canada, 277-280.
11. Chang, S. G., Bin, Yu. & Vetteli, M. (2000). Spatially adaptive wavelet thresholding with context modeling for image denoising. *IEEE Trans. Image Proc. 9*, 1522-1531.
12. Donoho, D. L. & Johnstone, I. M. (1994). Ideal spatial adaptation via waveletshrinkage. *Biometrika*, *81*, 425–455.
13. Donoho, D. L. & Johnstone, I. M. (1995). Adapting to unknown smoothness via wavelet shrinkage. *Journal of the American Statistical Assoc*, *90*(432), 1200–1224.
14. Donoho, D. L., Johnstone, I. M., Kerkyacharian, G. & Picard, D. (1995). Wavelet shrinkage: Asymptopia. *J. R. Stat. Soc. B*, *Ser. B, 57*(2), 301–369.
15. Stein, C. M. (1981). Estimation of the mean of a multivariate normal distribution. *Ann. Statist*, *9*(6), 1135–1151.
16. Jansen, M., Malfait, M. & Bultheel, A. (1997). Generalized cross-validation for wavelet thresholding. *Signal Processing*, *56*, 33–44.
17. Nason, G. P. (1996). Wavelet shrinkage using cross-validation. *J.R. Statist. Soc. B*, *58*, 463–479.
18. Weyrich, N. & Warhola, G. T. (1995a). De-noising using wavelets and cross-validation. *NATO Adv. Study Inst*. *C, 454*, 523–532.

19. Weyrich, N. & Warhola, G. T. (1995b). Wavelet shrinkage and generalized cross-validation for de-noising with applications to speech. In Wavelets and Multilevel Approximation, C.K. Chui, L.L. Schumaker, (Eds.), *Approximation Theory 8*(2), 407–414, Singapore: World Scientific.
20. Ogden, R. T. & Parzen, E. (1996a). Change-point approach to data analytic wavelet thresholding. *Statist. Comput*., *6*, 93–99.
21. Ogden, R. T. & Parzen, E. (1996b). Data dependent wavelet thresholding in nonparametric regression with change-point applications. *Comput. Statist. Data Anal.*, *22*, 53–70.
22. Abramovich, F. & Benjamini, Y. (1995). Thresholding of wavelet coefficients as multiple hypotheses testing procedure. In Wavelets and Statistics, Antoniadis, A. & Oppenheim, G. (Eds.), *Lecture Notes in Statistics 103*, 5–14, New York: Springer-Verlag.
23. Abramovich, F. & Benjamini, Y. (1996). Adaptive thresholding of wavelet coefficients. *Comput. Statist.*, *22*, 351–361.
24. Hall, P., Kerkyacharian, G. & Picard, D. (1998). Block threshold rules for curve estimation using kernel and wavelet methods. *Ann. Statist.*, *26*, 922–942.
25. Hall, P., Kerkyacharian, G. & Picard, D. (1999). On the minimax optimality of block thresholded wavelet estimators. Rules for curve estimation using kernel and wavelet methods. *Statist. Sinica.*, *9*, 33–50.
26. Hall, P., Penev, S., Kerkyacharian, G. & Picard, D. (1997). Numerical performance of block thresholded wavelet estimators, *Statist. Comput.*, *7*, 115–124.
27. Cai, T. T. & Brown, L. D. (1999). Wavelet estimation for samples with random uniform design. *Statist. Probab. Lett.*, *42*, 313–321.
28. Efromovich, S. (1999).Quasi-linear wavelet estimation. *Journal of the American Statistical Association, 94*(445), *94*, 189–204.
29. Efromovich, S. (2000). Sharp linear and block shrinkage wavelet estimation. *Statist. Probab.Lett.*, *49*.
30. Cai, T. T. & Silverman, B. W. (2001). Incorporating information on neighboring coefficients into wavelet estimation. *Sankhya, Series A*, *63*, 127-148.
31. Cai, T. T. (1999). Adaptive wavelet estimation: a block thresholding and oracle inequality approach. *Ann. Statist.*, *27*, 898–924.
32. DeVore, R. A. & Lucier, B. J. (1992). Fast wavelet techniques for near-optimal image processing. *in IEEE Military Communications Conf. Rec.* San Diego, *3*, 1129–1135.
33. Koza, J. R. (1990). Genetic programming: A paradigm for genetically breeding populations of computer programs to solve problems. Stanford University Computer Science Dept. *Technical Report STAN-CS*, *90*, 1314.
34. Harris, C. P. & Buxton, B. F. (1996). Evolving edge detectors with genetic programming. In Genetic Programming: *Proceeding of the First Conference.* Stanford University, CA, USA, MIT Press, 309-315.
35. Harris, C. & Buxton, B. (1997). Low-level edge detection using genetic programming: performance, specificity and application to real-world signals, *Research Note, Number RN/97/7,* UCL, Gover Street, London, WCIE 6BT, UK.
36. Mallat, S. (1999). A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Trans. Pattern Anal. Machine Intell., 11*, 674-693.
37. Mallat, S. (1989). Multifrequency channel decompositions of images and wavelet models, *IEEE Trans. Acous. Speech Signal Proc*., *37*(12), 2091-2110.
38. Holland, J. H. (1975). *Adaptation in Natural and Artificial Systems*, Ann Arbor, MI: University of Michigan Press.

39. Holland, J. H. & Burks, A. W. (1987). Adaptive computing system capable of learning and discovery, *U. S. Patent*, *4*, 697, 242.
40. Koza, J. R. & Keane, M. A. (1990). Genetic breeding of non-linear optimal control Strategies for broom balancing. *In Proceedings of the Ninth International Conference on Analysis and Optimization of Systems.* Berlin: Springer-Verlag, 203-232.
41. Koza, J. R. (1991). Genetic evolution and co-evolution of computer programs. In Langton, Christopher, Taylor, Charles, Farmer, J. Doyne, and Rasmussen, Steen (editors). *Artificial Life II, SFI Studies in the Sciences of Complexity*. Redwood City, CA: Addison-Wesley.*10*, 603-629.
42. Koza, J. R. (1992). *Genetic Programming*. Cambridge, MA: MIT Press.
43. Koza, J. R. (1991). *Evolution and co-evolution of computer programs to control independent acting agents*. In Meyer and Wilson below.