

Survey of Image Denoising Techniques

Mukesh C. Motwani
Image Process Technology, Inc.
1776 Back Country Road
Reno, NV 89521 USA
(775) 448-7816
mukesh@image-process.com

Mukesh C. Gadiya
University of Pune, India
Vishwakarma Inst. of Tech.
Pune 411337, INDIA
91-9884371488
mukesh_gadiya@satyam.com

Rakhi C. Motwani
University of Nevada, Reno
Dept of Comp. Sci. & Engr.,
Reno, NV 89557 USA
(775) 853-7897
rakhi@cs.unr.edu

Frederick C. Harris, Jr.
University of Nevada, Reno
Dept of Comp. Sci. & Engr.,
Reno, NV 89557 USA
(775) 784-6571
fredh@cs.unr.edu

Abstract

Removing noise from the original signal is still a challenging problem for researchers. There have been several published algorithms and each approach has its assumptions, advantages, and limitations. This paper presents a review of some significant work in the area of image denoising. After a brief introduction, some popular approaches are classified into different groups and an overview of various algorithms and analysis is provided. Insights and potential future trends in the area of denoising are also discussed.

1. Introduction

Digital images play an important role both in daily life applications such as satellite television, magnetic resonance imaging, computer tomography as well as in areas of research and technology such as geographical information systems and astronomy. Data sets collected by image sensors are generally contaminated by noise. Imperfect instruments, problems with the data acquisition process, and interfering natural phenomena can all degrade the data of interest. Furthermore, noise can be introduced by transmission errors and compression. Thus, denoising is often a necessary and the first step to be taken before the images data is analyzed. It is necessary to apply an efficient denoising technique to compensate for such data corruption.

Image denoising still remains a challenge for researchers because noise removal introduces artifacts and causes blurring of the images. This paper

describes different methodologies for noise reduction (or denoising) giving an insight as to which algorithm should be used to find the most reliable estimate of the original image data given its degraded version.

Noise modeling in images is greatly affected by capturing instruments, data transmission media, image quantization and discrete sources of radiation. Different algorithms are used depending on the noise model. Most of the natural images are assumed to have additive random noise which is modeled as a Gaussian. Speckle noise [1] is observed in ultrasound images whereas Rician noise [2] affects MRI images. The scope of the paper is to focus on noise removal techniques for natural images.

2. Evolution of Image Denoising Research

Image Denoising has remained a fundamental problem in the field of image processing. Wavelets give a superior performance in image denoising due to properties such as sparsity and multiresolution structure. With Wavelet Transform gaining popularity in the last two decades various algorithms for denoising in wavelet domain were introduced. The focus was shifted from the Spatial and Fourier domain to the Wavelet transform domain. Ever since Donoho's Wavelet based thresholding approach was published in 1995, there was a surge in the denoising papers being published. Although Donoho's concept was not revolutionary, his methods did not require tracking or correlation of the wavelet maxima and minima across the different scales as proposed by Mallat [3]. Thus, there was a renewed interest in wavelet based denoising techniques since Donoho [4] demonstrated a simple approach to a difficult problem. Researchers

published different ways to compute the parameters for the thresholding of wavelet coefficients. Data adaptive thresholds [6] were introduced to achieve optimum value of threshold. Later efforts found that substantial improvements in perceptual quality could be obtained by translation invariant methods based on thresholding of an Undecimated Wavelet Transform [7]. These thresholding techniques were applied to the non-orthogonal wavelet coefficients to reduce artifacts. Multiwavelets were also used to achieve similar results. Probabilistic models using the statistical properties of the wavelet coefficient seemed to outperform the thresholding techniques and gained ground. Recently, much effort has been devoted to Bayesian denoising in Wavelet domain. Hidden Markov Models and Gaussian Scale Mixtures have also become popular and more research continues to be published. Tree Structures ordering the wavelet coefficients based on their magnitude, scale and spatial location have been researched. Data adaptive transforms such as Independent Component Analysis (ICA) have been explored for sparse shrinkage. The trend continues to focus on using different statistical models to model the statistical properties of the wavelet coefficients and its neighbors. Future trend will be towards finding more accurate probabilistic models for the distribution of non-orthogonal wavelet coefficients.

3. Classification of Denoising Algorithms

As shown in Figure 1, there are two basic approaches to image denoising, spatial filtering methods and transform domain filtering methods.

3.1 Spatial Filtering

A traditional way to remove noise from image data is to employ spatial filters. Spatial filters can be further classified into non-linear and linear filters.

I. Non-Linear Filters

With non-linear filters, the noise is removed without any attempts to explicitly identify it. Spatial filters employ a low pass filtering on groups of pixels with the assumption that the noise occupies the higher region of frequency spectrum. Generally spatial filters remove noise to a reasonable extent but at the cost of blurring images which in turn makes the edges in pictures invisible. In recent years, a variety of nonlinear median-type filters such as weighted median [8], rank conditioned rank selection [9], and relaxed median [10] have been developed to overcome this drawback.

II. Linear Filters

A mean filter is the optimal linear filter for Gaussian noise in the sense of mean square error. Linear filters too tend to blur sharp edges, destroy lines and other fine image details, and perform poorly in the presence of signal-dependent noise. The Wiener filtering [11] method requires the information about the spectra of the noise and the original signal and it works well only if the underlying signal is smooth. Wiener method implements spatial smoothing and its model complexity control correspond to choosing the window size. To overcome the weakness of the Wiener filtering, Donoho and Johnstone proposed the wavelet based denoising scheme in [12, 13].

3.2 Transform Domain Filtering

The transform domain filtering methods can be subdivided according to the choice of the basis functions. The basis functions can be further classified as data adaptive and non-adaptive. Non-adaptive transforms are discussed first since they are more popular.

3.2.1 Spatial-Frequency Filtering

Spatial-frequency filtering refers use of low pass filters using Fast Fourier Transform (FFT). In frequency smoothing methods [11] the removal of the noise is achieved by designing a frequency domain filter and adapting a cut-off frequency when the noise components are decorrelated from the useful signal in the frequency domain. These methods are time consuming and depend on the cut-off frequency and the filter function behavior. Furthermore, they may produce artificial frequencies in the processed image.

3.2.2 Wavelet domain

Filtering operations in the wavelet domain can be subdivided into linear and nonlinear methods.

I. Linear Filters

Linear filters such as Wiener filter in the wavelet domain yield optimal results when the signal corruption can be modeled as a Gaussian process and the accuracy criterion is the mean square error (MSE) [14, 15]. However, designing a filter based on this assumption frequently results in a filtered image that is more visually displeasing than the original noisy signal, even though the filtering operation successfully reduces the MSE. In [16] a wavelet-domain spatially-adaptive FIR Wiener filtering for image denoising is proposed where Wiener filtering is performed only within each scale and intrascale filtering is not allowed.

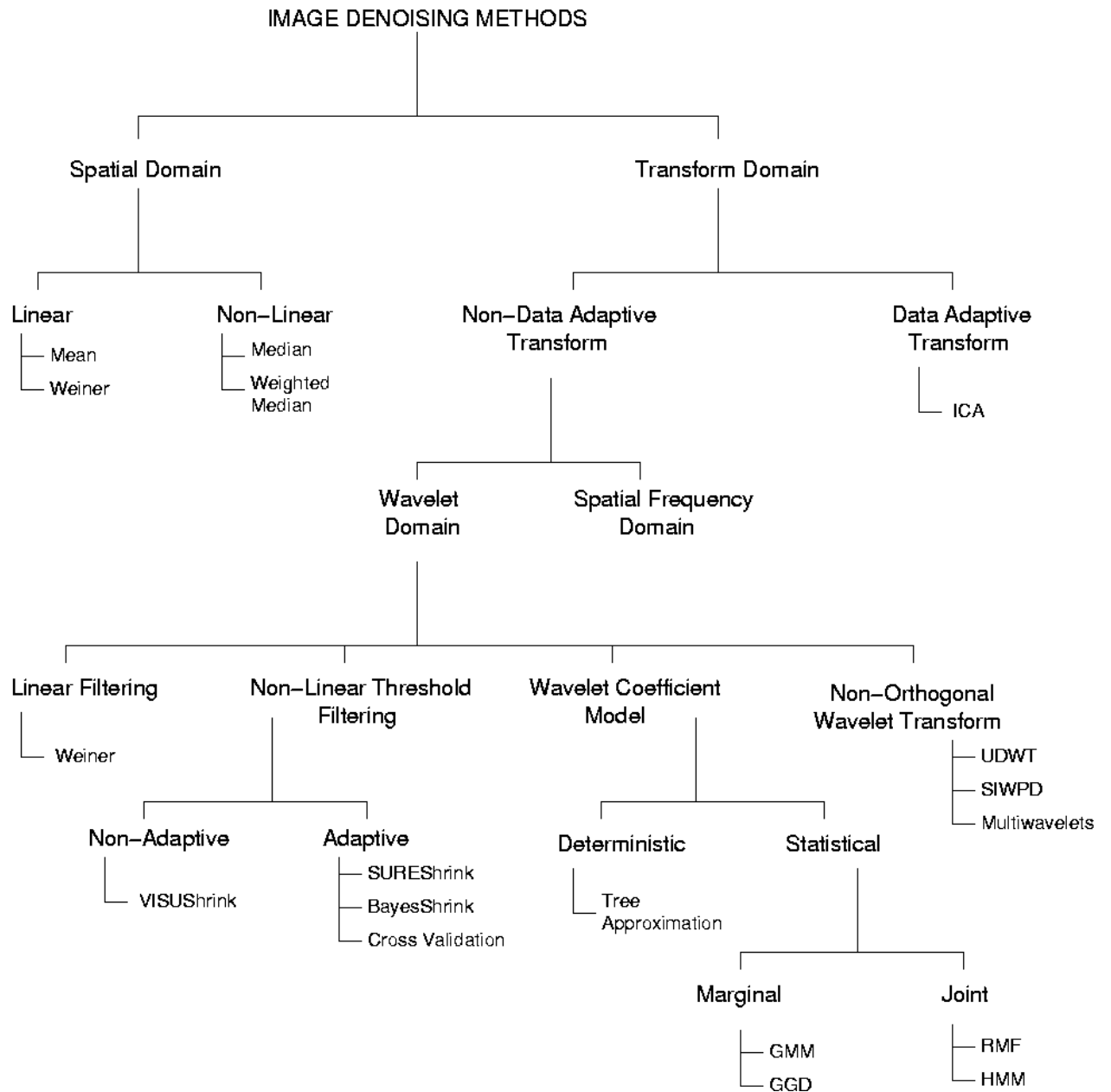


Figure 1 – Classification of Image Denoising Methods

II. Non-Linear Threshold Filtering

The most investigated domain in denoising using Wavelet Transform is the non-linear coefficient thresholding based methods. The procedure exploits sparsity property of the wavelet transform and the fact that the Wavelet Transform maps white noise in the signal domain to white noise in the transform domain. Thus, while signal energy becomes more concentrated into fewer coefficients in the transform domain, noise energy does not. It is this important principle that enables the separation of signal from noise.

The procedure in which small coefficients are removed while others are left untouched is called Hard Thresholding [5]. But the method generates spurious blips, better known as artifacts, in the images as a result of unsuccessful attempts of removing moderately large noise coefficients. To overcome the demerits of hard thresholding, wavelet transform using soft thresholding was also introduced in [5]. In this scheme, coefficients above the threshold are shrunk by the absolute value of the threshold itself. Similar to soft thresholding, other techniques of applying thresholds are semi-soft thresholding and Garrote thresholding [6]. Most of the wavelet shrinkage literature is based on methods for choosing the optimal threshold which can be adaptive or non-adaptive to the image.

a. Non-Adaptive thresholds

VISUShrink [12] is non-adaptive universal threshold, which depends only on number of data points. It has asymptotic equivalence suggesting best performance in terms of MSE when the number of pixels reaches infinity. VISUShrink is known to yield overly smoothed images because its threshold choice can be unwarrantedly large due to its dependence on the number of pixels in the image.

b. Adaptive Thresholds

SUREShrink [12] uses a hybrid of the universal threshold and the SURE [Stein's Unbiased Risk Estimator] threshold and performs better than VISUShrink. BayesShrink [17, 18] minimizes the Bayes' Risk Estimator function assuming Generalized Gaussian prior and thus yielding data adaptive threshold. BayesShrink outperforms SUREShrink most of the times. Cross Validation [19] replaces wavelet coefficient with the weighted average of neighborhood coefficients to minimize generalized cross validation (GCV) function providing optimum threshold for every coefficient.

The assumption that one can distinguish noise from the signal solely based on coefficient magnitudes is violated when noise levels are higher than signal magnitudes. Under this high noise circumstance, the spatial configuration of neighboring wavelet coefficients

can play an important role in noise-signal classifications. Signals tend to form meaningful features (e.g. straight lines, curves), while noisy coefficients often scatter randomly.

III. Non-orthogonal Wavelet Transforms

Undecimated Wavelet Transform (UDWT) has also been used for decomposing the signal to provide visually better solution. Since UDWT is shift invariant it avoids visual artifacts such as pseudo-Gibbs phenomenon. Though the improvement in results is much higher, use of UDWT adds a large overhead of computations thus making it less feasible. In [20] normal hard/soft thresholding was extended to Shift Invariant Discrete Wavelet Transform. In [21] Shift Invariant Wavelet Packet Decomposition (SIWPD) is exploited to obtain number of basis functions. Then using Minimum Description Length principle the Best Basis Function was found out which yielded smallest code length required for description of the given data. Then, thresholding was applied to denoise the data. In addition to UDWT, use of Multiwavelets is explored which further enhances the performance but further increases the computation complexity. The Multiwavelets are obtained by applying more than one mother function (scaling function) to given dataset. Multiwavelets possess properties such as short support, symmetry, and the most importantly higher order of vanishing moments. This combination of shift invariance & Multiwavelets is implemented in [22] which give superior results for the Lena image in context of MSE.

IV. Wavelet Coefficient Model

This approach focuses on exploiting the multiresolution properties of Wavelet Transform. This technique identifies close correlation of signal at different resolutions by observing the signal across multiple resolutions. This method produces excellent output but is computationally much more complex and expensive. The modeling of the wavelet coefficients can either be deterministic or statistical.

a. Deterministic

The Deterministic method of modeling involves creating tree structure of wavelet coefficients with every level in the tree representing each scale of transformation and nodes representing the wavelet coefficients. This approach is adopted in [23]. The optimal tree approximation displays a hierarchical interpretation of wavelet decomposition. Wavelet coefficients of singularities have large wavelet coefficients that persist along the branches of tree. Thus if a wavelet coefficient has strong presence at particular node then in case of it being signal, its presence should be more pronounced at its parent nodes. If it is noisy coefficient, for instance spurious

blip, then such consistent presence will be missing. Lu et al. [24], tracked wavelet local maxima in scale-space, by using a tree structure. Other denoising method based on wavelet coefficient trees is proposed by Donoho [25].

b. Statistical Modeling of Wavelet Coefficients

This approach focuses on some more interesting and appealing properties of the Wavelet Transform such as multiscale correlation between the wavelet coefficients, local correlation between neighborhood coefficients etc. This approach has an inherent goal of perfecting the exact modeling of image data with use of Wavelet Transform. A good review of statistical properties of wavelet coefficients can be found in [26] and [27]. The following two techniques exploit the statistical properties of the wavelet coefficients based on a probabilistic model.

i. Marginal Probabilistic Model

A number of researchers have developed homogeneous local probability models for images in the wavelet domain. Specifically, the marginal distributions of wavelet coefficients are highly kurtotic, and usually have a marked peak at zero and heavy tails. The Gaussian mixture model (GMM) [28] and the generalized Gaussian distribution (GGD) [29] are commonly used to model the wavelet coefficients distribution. Although GGD is more accurate, GMM is simpler to use. In [30], authors proposed a methodology in which the wavelet coefficients are assumed to be conditionally independent zero-mean Gaussian random variables, with variances modeled as identically distributed, highly correlated random variables. An approximate Maximum A Posteriori (MAP) Probability rule is used to estimate marginal prior distribution of wavelet coefficient variances. All these methods mentioned above require a noise estimate, which may be difficult to obtain in practical applications. Simoncelli and Adelson [33] used a two-parameter generalized Laplacian distribution for the wavelet coefficients of the image, which is estimated from the noisy observations. Chang et al. [34] proposed the use of adaptive wavelet thresholding for image denoising, by modeling the wavelet coefficients as a generalized Gaussian random variable, whose parameters are estimated locally (i.e., within a given neighborhood).

ii. Joint Probabilistic Model

Hidden Markov Models (HMM) [35] models are efficient in capturing inter-scale dependencies, whereas Random Markov Field [36] models are more efficient to capture intrascale correlations. The complexity of local structures is not well described by Random Markov Gaussian densities whereas Hidden Markov Models can be used to capture higher order statistics. The

correlation between coefficients at same scale but residing in a close neighborhood are modeled by Hidden Markov Chain Model where as the correlation between coefficients across the chain is modeled by Hidden Markov Trees. Once the correlation is captured by HMM, Expectation Maximization is used to estimate the required parameters and from those, denoised signal is estimated from noisy observation using well-known MAP estimator. In [31], a model is described in which each neighborhood of wavelet coefficients is described as a Gaussian scale mixture (GSM) which is a product of a Gaussian random vector, and an independent hidden random scalar multiplier. Strela et al. [32] described the joint densities of clusters of wavelet coefficients as a Gaussian scale mixture, and developed a maximum likelihood solution for estimating relevant wavelet coefficients from the noisy observations. Another approach that uses a Markov random field model for wavelet coefficients was proposed by Jansen and Bulthel [37]. A disadvantage of HMT is the computational burden of the training stage. In order to overcome this computational problem, a simplified HMT, named as uHMT [27], was proposed.

3.2.3 Data-Adaptive Transforms

Recently a new method called Independent Component Analysis (ICA) has gained wide spread attention. The ICA method was successfully implemented in [38, 39] in denoising Non-Gaussian data. One exceptional merit of using ICA is it's assumption of signal to be Non-Gaussian which helps to denoise images with Non-Gaussian as well as Gaussian distribution. Drawbacks of ICA based methods as compared to wavelet based methods are the computational cost because it uses a sliding window and it requires sample of noise free data or at least two image frames of the same scene. In some applications, it might be difficult to obtain the noise free training data.

4. Discussion

Performance of denoising algorithms is measured using quantitative performance measures such as peak signal-to-noise ratio (PSNR), signal-to-noise ratio (SNR) as well as in terms of visual quality of the images. Many of the current techniques assume the noise model to be Gaussian. In reality, this assumption may not always hold true due to the varied nature and sources of noise. An ideal denoising procedure requires *a priori* knowledge of the noise, whereas a practical procedure may not have the required information about the variance of the noise or the noise

model. Thus, most of the algorithms assume known variance of the noise and the noise model to compare the performance with different algorithms. Gaussian Noise with different variance values is added in the natural images to test the performance of the algorithm. Not all researchers use high value of variance to test the performance of the algorithm when the noise is comparable to the signal strength.

Use of FFT in filtering has been restricted due to its limitations in providing sparse representation of data. Wavelet Transform is the best suited for performance because of its properties like sparsity, multiresolution and multiscale nature. In addition to performance, issues of computational complexity must also be considered. Thresholding techniques used with the Discrete Wavelet Transform are the simplest to implement. Non-orthogonal wavelets such as UDWT and Multiwavelets improve the performance at the expense of a large overhead in their computation. HMM based methods seem to be promising but are complex.

When using Wavelet Transform, Nason [40] emphasized that issue such as choice of primary resolution (the scale level at which to begin thresholding) and choice of analyzing wavelet also have a large influence on the success of the shrinkage procedure. When comparing algorithms, it is very important that researchers do not omit these comparison details. Several papers did not specify the wavelet used neither the level of decomposition of the wavelet transform was mentioned.

It is expected that the future research will focus on building robust statistical models of non-orthogonal wavelet coefficients based on their intra scale and inter scale correlations. Such models can be effectively used for image denoising and compression.

5. References

[1] H. Guo, J. E. Odegard, M. Lang, R. A. Gopinath, I. W. Selesnick, and C. S. Burrus, "Wavelet based speckle reduction with application to SAR based ATD/R," First Int'l Conf. on Image Processing, vol. 1, pp. 75-79, Nov. 1994.

[2] Robert D. Nowak, "Wavelet Based Rician Noise Removal", IEEE Transactions on Image Processing, vol. 8, no. 10, pp.1408, October 1999.

[4] S. G. Mallat and W. L. Hwang, "Singularity detection and processing with wavelets," IEEE Trans. Inform. Theory, vol. 38, pp. 617-643, Mar. 1992.

[5] D. L. Donoho, "De-noising by soft-thresholding", IEEE Trans. Information Theory, vol.41, no.3, pp.613-

627, May1995. <http://www-stat.stanford.edu/~donoho/Reports/1992/denoisereleas e3.ps.Z>

[6] Imola K. Fodor, Chandrika Kamath, "Denoising through wavelet shrinkage: An empirical study", Center for applied science computing Lawrence Livermore National Laboratory, July 27, 2001.

[7] R. Coifman and D. Donoho, "Translation invariant de-noising," in Lecture Notes in Statistics: Wavelets and Statistics, vol. New York: Springer-Verlag, pp. 125-150, 1995.

[8] R. Yang, L. Yin, M. Gabbouj, J. Astola, and Y. Neuvo, "Optimal weighted median filters under structural constraints," IEEE Trans. Signal Processing, vol. 43, pp. 591-604, Mar. 1995.

[9] R. C. Hardie and K. E. Barner, "Rank conditioned rank selection filters for signal restoration," IEEE Trans. Image Processing, vol. 3, pp.192-206, Mar. 1994.

[10] A. Ben Hamza, P. Luque, J. Martinez, and R. Roman, "Removing noise and preserving details with relaxed median filters," J. Math. Imag. Vision, vol. 11, no. 2, pp. 161-177, Oct. 1999.

[11] A.K.Jain, Fundamentals of digital image processing. Prentice-Hall, 1989

[12] David L. Donoho and Iain M. Johnstone, "Ideal spatial adaption via wavelet shrinkage", Biometrika, vol.81, pp 425-455, September 1994.

[13] David L. Donoho and Iain M. Johnstone., "Adapting to unknown smoothness via wavelet shrinkage", Journal of the American Statistical Association, vol.90, no432, pp.1200-1224, December 1995. National Laboratory, July 27, 2001.

[14] V. Strela. "Denoising via block Wiener filtering in wavelet domain". In 3rd European Congress of Mathematics, Barcelona, July 2000. Birkhäuser Verlag.

[15] H. Choi and R. G. Baraniuk, "Analysis of wavelet domain Wiener filters," in IEEE Int. Symp. Time-Frequency and Time-Scale Analysis, (Pittsburgh), Oct. 1998. <http://citeseer.ist.psu.edu/article/choi98analysis.html>

[16] H. Zhang, Aria Nosratinia, and R. O. Wells, Jr., "Image denoising via wavelet-domain spatially adaptive FIR Wiener filtering", in IEEE Proc. Int. Conf. Acoust., Speech, Signal Processing, Istanbul, Turkey, June 2000.

[17] E. P. Simoncelli and E. H. Adelson. Noise removal via Bayesian wavelet coring. In Third Int'l Conf on Image Proc, volume I, pages 379-382, Lausanne, September 1996. IEEE Signal Proc Society.

[18]. H. A. Chipman, E. D. Kolaczyk, and R. E. McCulloch: 'Adaptive Bayesian wavelet shrinkage', J. Amer. Stat. Assoc., Vol. 92, No 440, Dec. 1997, pp. 1413-1421.

[19] Marteen Jansen, Ph. D. Thesis in "Wavelet thresholding and noise reduction" 2000.

[20] M. Lang, H. Guo, J.E. Odegard, and C.S. Burrus, "Nonlinear processing of a shift invariant DWT for

noise reduction," SPIE, Mathematical Imaging: Wavelet Applications for Dual Use, April 1995.

[21] I. Cohen, S. Raz and D. Malah, Translation-invariant denoising using the minimum description length criterion, *Signal Processing*, 75, 3, 201-223, (1999).

[22] T. D. Bui and G. Y. Chen, "Translation-invariant denoising using multiwavelets", *IEEE Transactions on Signal Processing*, Vol.46, No.12, pp.3414-3420, 1998.

[23] R. G. Baraniuk, "Optimal tree approximation with wavelets," in *Proc. SPIE Tech. Conf. Wavelet Applications Signal Processing VII*, vol. 3813, Denver, CO, 1999, pp. 196-207.

[24] J. Lu, J. B. Weaver, D.M. Healy, and Y. Xu, "Noise reduction with multiscale edge representation and perceptual criteria," in *Proc. IEEE-SP Int. Symp. Time-Frequency and Time-Scale Analysis*, Victoria, BC, Oct. 1992, pp. 555-558.

[25] D. L. Donoho, "CART and best-ortho-basis: A connection," *Ann. Statist.*, pp. 1870-1911, 1997.

[26] R. W. Buccigrossi, and E. P. Simoncelli, 'Image compression via joint statistical characterization in the wavelet domain', *IEEE Image Process.*, Vol. 8, No 12, Dec.1999, pp. 1688-1701.

[27] J. K. Romberg, H. Choi, and R. G. Baraniuk, "Bayesian tree-structured image modeling using wavelet-domain hidden Markov models", *IEEE Image Process.*, Vol. 10, No 7, Jul. 2001, pp. 1056-1068.

[28] H. A. Chipman, E. D. Kolaczyk, and R. E. McCulloch: "Adaptive Bayesian wavelet shrinkage", *J. Amer. Stat. Assoc.*, Vol. 92, No 440, Dec. 1997, pp. 1413-1421.

[29] P. Moulin and J. Liu, "Analysis of multiresolution image denoising schemes using generalized Gaussian and complexity priors", *IEEE Infor. Theory*, Vol. 45, No 3, Apr. 1999, pp. 909-919.

[30] M. K. Mihcak, I. Kozintsev, K. Ramchandran, and P. Moulin, "Low-Complexity Image Denoising Based on Statistical Modeling of Wavelet Coefficients," *IEEE Signal Processing Lett.* (to appear).

[31] Portilla, J., Strela, V., Wainwright, M., Simoncelli, E.P., "Image Denoising using Gaussian Scale Mixtures

in the Wavelet Domain", TR2002-831, Computer Science Dept, New York University. 2002.

[32] V. Strela, J. Portilla, and E. P. Simoncelli, "Image denoising via a local Gaussian scale mixture model in the wavelet domain," in *Proc. SPIE 45th Annu. Meeting*, San Diego, CA, Aug. 2000.

[33] E. P. Simoncelli and E. Adelson, "Noise removal via Bayesian wavelet coring," in *Proc. IEEE International Conference on Image Processing*, Lausanne, Switzerland, September 1996, pp. 279-382.

[34] S. G. Chang, B. Yu, and M. Vetterli, "Spatially adaptive wavelet thresholding with context modeling for image denoising," *IEEE Trans. Image Processing*, vol. 9, pp. 1522-1531, Sept. 2000.

[35] J. Romberg, H. Choi and R. G. Baraniuk, "Bayesian wavelet domain image modeling using hidden Markov models," *IEEE Transactions on Image Processing*, vol. 10, pp. 1056-1068, July 2001.

[36] M. Malfait and D. Roose, "Wavelet based image denoising using a Markov Random Field *a priori* model," *IEEE Transactions on Image Processing*, vol. 6, no. 4, pp. 549-565, 1997.

[37] M. Jansen and A. Bulthel, "Empirical bayes approach to improve wavelet thresholding for image noise reduction," *Journal of the American Statistical Association*, vol. 96, no. 454, pp. 629-639, June 2001.

[38] A. Jung, "An introduction to a new data analysis tool: Independent Component Analysis", *Proceedings of Workshop GK "Nonlinearity" - Regensburg*, Oct. 2001.

[39] A. Hyvärinen, E. Oja, P. Hoyer, and J. Hurri, "Image feature extraction by sparse coding and independent component analysis", In *Proc. Int. Conf. on Pattern Recognition (ICPR'98)*, pp. 1268-1273, Brisbane, Australia, 1998.

[40] Nason, G.P., "Fast cross-validatory choice of wavelet smoothness, primary resolution and threshold in wavelet shrinkage using the Kovac-Silverman algorithm", Technical report, Department of Mathematics, University of Bristol, United Kingdom.