ON THE NOISE REDUCTION USING WAVELET TRANSFORMS

Tertulien Ndjountche and Rolf Unbehauen*

Lehrstuhl für Allgemeine und Theoretische Elektrotechnik
Universität Erlangen–Nürnberg
Cauerstraße 7, 91058 Erlangen
Germany
(*) E–mail: tertu@late.e–technik.uni–erlangen.de

Abstract - ***Wavelet–based techniques are more efficient for recovering a signal corrupted by noise. The time– and frequency–localization capabilities of wavelets provide better noise reduction and less signal distortion than conventional filtering methods. The noise reduction technique used in this paper is based on the hidden Markov tree (HMT) structure, which can efficiently model the statistical characteristics of practical signals. As confirmed by numerical results, the HMT based approach provides a significant performance improvement over competing methods.***

Keywords: wavelet transform, thresholding, hidden Markov tree model, expectation–maximization algorithm, noise reduction.

1. INTRODUCTION

Wavelet transforms appear as an efficient tool for signal and image processing tasks, which can involve real–world data. In particular, they can localize the signal information in the time–frequency plane with different resolutions. This property has been found to be useful in noise reduction [1, 2, 3].

In [2, 3], a procedure is proposed for recovering functions from noisy data through the application of the following steps.

- Discretize the continuous function $f(t)$ into $f(i)$, $0 \leq i \leq N-1$, via sampling.
- Transform the sequence $f(i)$, $0 \leq i \leq N-1$, into an orthogonal domain by the discrete wavelet transform. The resulting coefficients will represent the function information on various scales.
- Apply soft or hard thresholding to the wavelet coefficients; the threshold may be scale dependent or not.
- Perform the inverse discrete wavelet transform to restore the de–noised function.

This algorithm for the noise reduction is based on the principle that the wavelet coefficients below a threshold are discarded and only the larger ones corresponding to signal features are retained (hard thresholding) or the threshold is subtracted from all the coefficients (soft thresholding). It offers the advantages of smoothness (i.e., the de–noised estimate has a very high probability of being as smooth as the original signal in a variety of spaces) and adaptation.

Thresholding methods have demonstrated near–ideal estimation performances in various asymptotic frameworks. However, they can be outperformed in practice by Bayesian techniques, which assume the independence of the wavelet coefficients. The hidden Markov tree (HMT) structures adopted for the wavelet transform description in this paper match the statistical characteristics of non Gaussian data. Their parameters are recursively estimated until the convergence. The HMT models are sufficiently flexible to model representations of processes in arbitrary wavelet function bases and thereby improve the results of the signal detection or restoration.

2. WAVELET BACKGROUND

In multiresolution analysis, the signal is approximated at various resolutions with orthogonal projections on different spaces, $\{V_j\}_{j\in\mathbb{Z}}$, which are nested such that $V_{j+1} \subset V_j$. Furthermore, it is required to have

$$\overline{\bigcup_{j\in\mathbb{Z}} V_j} = L_2(\mathbb{R}) \qquad (1)$$

This work was supported in part by the German Research Society DFG.

$$\bigcap_{j\in\mathbb{Z}} V_j = \{0\} \tag{2}$$

$$\forall (j,k) \in \mathbb{Z}^2, f(t) \in V_j \Leftrightarrow f(t-2^j k) \in V_j \tag{3}$$

and

$$\forall j \in \mathbb{Z}, f(t) \in V_j \Leftrightarrow f(2t) \in V_{j+1}, \tag{4}$$

where $L_2(\mathbb{R})$ is the space of square integrable functions. Let us assume that the finest scale of interest is described by the subspace V_0 and consider only its circulant shifts $\{\phi_{j_0,k}\}_{k\in\mathbb{Z}}$. Here, $\phi_{j_0,k}$ denotes a translation and dilation of the scaling function ϕ, i.e.,

$$\phi_{j_0,k}(t) = 2^{-j_0/2}\phi(2^{-j_0}t - k). \tag{5}$$

The projection of a function $f \in L_2(\mathbb{R})$ onto V_0 can be defined by the scalar product

$$p_{j_0,k} = \langle f, \phi_{j_0,k}\rangle. \tag{6}$$

Assuming that W_j is the orthonormal complement of V_j in V_{j-1}, we can write

$$V_{j-1} = V_j \oplus W_j. \tag{7}$$

This subspace is generated by an orthonormal wavelet basis, $\{\psi_{j,k}\}_{(j,k)\in\mathbb{Z}^2}$, where

$$\psi_{j,k}(t) = 2^{-j/2}\psi(2^{-j}t - k). \tag{8}$$

The scalar product resulting from the projection of a function f onto W_j is defined as

$$q_{j,k} = \langle f, \psi_{j,k}\rangle. \tag{9}$$

The wavelet decomposition of a signal, f, can be given by

$$f = \sum_{j\in\mathbb{Z}}\sum_{k\in\mathbb{Z}} \langle f, \psi_{j,k}\rangle \psi_{j,k}, \tag{10}$$

where the inner product $\langle f, \psi_{j,k}\rangle$ denotes the wavelet coefficients $q_{j,k}$ of f. The next mixed expansion can also be formed

$$f = \sum_{k\in\mathbb{Z}} \langle f, \phi_{j_0,k}\rangle \phi_{j_0,k} + \sum_{j\geq j_0}\sum_{k\in\mathbb{Z}} \langle f, \psi_{j,k}\rangle \psi_{j,k}, \tag{11}$$

where j_0 represents the coarsest resolution.

Given a discrete set of data $f(i)$, there exists an orthogonal matrix $\mathcal{W}$ such that the discrete wavelet transform (DWT) is given by

$$\mathbf{d} = \mathcal{W}\mathbf{f}, \tag{12}$$

where $\mathbf{f}$ and $\mathbf{d}$ are the vector of data and discrete wavelet coefficients, d_{jk}, respectively. In practice, the DWT is performed using an efficient algorithm or a multirate filterbank that only requires $O(N)$ operations. The spectral component of the signal existing at a given interval of time can be deduced from the variables $d_{j,k}$. The wavelet transform can then be interpreted as a frequency decomposition with a spatial orientation.

3. NOISE REDUCTION BY THRESHOLDING

Let $y(i) = x(i) + \sigma z(i)$, $0 \leq i \leq N-1$, be a finite length function of the signal observations. The signal of interest $x(i)$ is corrupted by a noise with the power σ^2 and $z(i)$ is a unit–variance, zero mean, Gaussian white noise, i.e.[1], $z \sim \mathcal{N}(0,1)$. For convenience of notation, the following vector definitions have to be introduced: $\mathbf{x} = [x(0), x(1), \ldots, x(N-1)]^T$, $\mathbf{y} = [y(0), y(1), \ldots, y(N-1)]^T$ and $\mathbf{z} = [z(0), z(1), \ldots, z(N-1)]^T$. The aim is to find an estimate $\hat{\mathbf{x}}$ of $\mathbf{x}$ from the noisy observation $\mathbf{y}$ with a small mean–squared error,

$$\xi = E||\hat{\mathbf{x}} - \mathbf{x}||^2, \tag{13}$$

where $||\cdot||$ denotes the Euclidean norm. Assuming that $\mathcal{W}$ denotes the DWT matrix, the DWT of $\mathbf{y}$ is given by

$$\mathbf{Y} = \mathcal{W}\mathbf{y} = \mathbf{y}^x + \sigma\mathbf{y}^z \tag{14}$$

and, inversely, we have

$$\mathbf{y} = \mathcal{W}^T\mathbf{Y} = \mathcal{W}^T\mathbf{y}^x + \sigma\mathcal{W}^T\mathbf{y}^z, \tag{15}$$

where the components of $\mathbf{Y}$ are indexed dyadically, i.e., $\mathbf{Y} = [Y_{j,k}]_{j\geq j_0;\ 0\leq k\leq 2^j-1}$, $\mathbf{y}^x$ and $\mathbf{y}^z$ are the wavelet coefficient vectors of the signal $\mathbf{x}$ and the noise $\mathbf{z}$, respectively. Note that since the transformation is orthogonal, $\mathbf{y}^z$ and $\mathbf{z}$ have the same nature and $\xi = E||\mathbf{y}^{\hat{x}} - \mathbf{y}^x||^2$, where $\mathbf{y}^{\hat{x}}$ the DWT of $\hat{\mathbf{x}}$, which can be computed as,

$$\hat{\mathbf{x}} = \mathcal{W}^T\mathbf{y}^{\hat{x}} = \mathcal{W}^T\Delta\mathbf{Y}, \tag{16}$$

where $\Delta = \text{diag}(\delta_0, \delta_1, \ldots, \delta_{N-1})$, $\delta_l \in \{1, 0\}$, represents a linear projection. It was proposed in [2] to derive the elements, $y^{\hat{x}}_{j,k}$, of $\mathbf{y}^{\hat{x}}$ by using either the soft threshold function given by

$$y^{\hat{x}}_{j,k} = \begin{cases} \text{sgn}(Y_{j,k})(|Y_{j,k}| - T), & \text{if } |Y_{j,k}| \geq T \\ 0 & \text{otherwise} \end{cases} \tag{17}$$

[1]The notation $x \sim \mathcal{N}(m_x, C_x)$ denotes that x is a normal random vector with mean m_x and covariance C_x.

or the hard threshold one of the form

$$y_{j,k}^{\hat{x}} = \begin{cases} Y_{j,k}, & \text{if } |Y_{j,k}| \geq T \\ 0, & \text{otherwise.} \end{cases} \quad (18)$$

Based on the minimax principle, the threshold, T, can be derived as $T = \sigma\sqrt{2\log(N)}$, where N is the number of data samples. This choice is simple but is suitable only for the cases of a signal corrupted by an independent and identically distributed Gaussian noise. Various alternatives to the minimax scheme [4] can be considered. The use of such methods makes it possible to select a scale dependent threshold in order to improve the estimation performance.

Generally, the true value of the noise standard deviation, σ, is unknown and it is often replaced by $\hat{\sigma} = \text{MAD}/0.6745$, where MAD is the median absolute value of the finest scale wavelet coefficients [2]. The restoration of data can be performed by the wavelet thresholding structure depicted in Fig. 1.

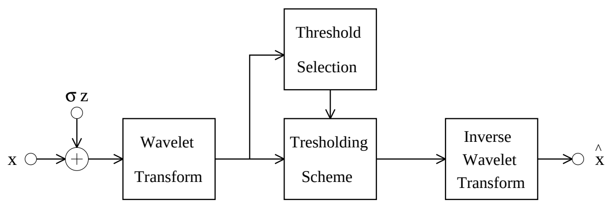

Figure 1: Wavelet thresholding building block for the noise reduction.

4. HMT WAVELET TRANSFORM

Assuming that the wavelet coefficients are divided into M classes, each wavelet coefficient can be associated to a discrete hidden state S_i, which can take on the values $m \in \{1, 2, \dots, M\}$. The state S_i corresponds to a Gaussian pdf with the variance μ_i. The overall pdf of a wavelet coefficient, ω_i, is given by

$$f(\omega_i) = \sum_{m=1}^{M} p_{S_i}(m) f(\omega_i | S_i = m), \quad (19)$$

where

$$f(\omega_i | S_i = m) = \frac{1}{\sqrt{2\pi}\sigma_{S_i}} \exp\left[-\frac{(\omega_i - \mu_i)^2}{2\sigma_{S_i}^2}\right], \quad (20)$$

and

$$\sum_{m=1}^{M} p_{S_i}(m) = 1. \quad (21)$$

The parameters $p_{S_i}(m)$ denotes the probability that ω_i belongs to a given class i or the probability mass function (pmf). Note that the specific case of the two–state model with the zero–mean Gaussian pdf is simple and useful for some applications.

Probabilistic graphs can be adopted to describe the local dependencies between the wavelet coefficients. These latter are associated to the nodes of the graph and dependencies between a pair of coefficients are illustrated by the connection between the corresponding nodes. Due to the computational complexity of a more general graph [5] and the properties of wavelet transform (locality and multiresolution), the special cases of hidden Markov based Models [6, 7, 8] appear as a suitable compromise.

Fig. 2 shows the multiscale structure associated to the wavelet coefficients. The random field of coefficients is segmented into classes of distinct statistical behavior. A tree structured dependency is assumed between the state variables, which are connected vertically from scale to scale. In the case of a three scale wavelet structure, this leads to the HMT model shown in Fig. 3.

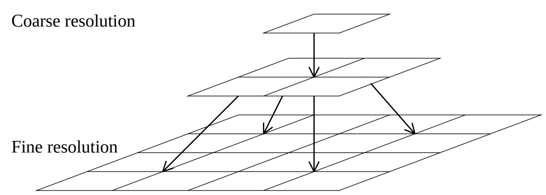

Figure 2: Structure of multiscale random wavelet coefficients.

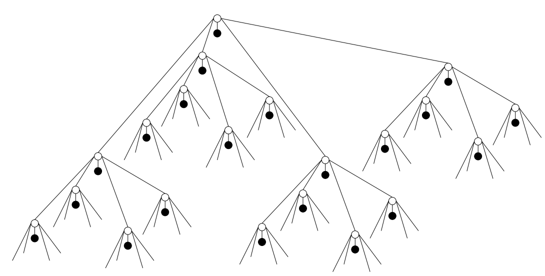

Figure 3: HMT structure.
The wavelet coefficient, ω_i, and its state variable, S_i, are associated to the black and white nodes, respectively.

Let the set of parameters $\boldsymbol{\theta}$ corresponding to the HMT be composed of the means, variances and probabilities of the states. The wavelet coefficients are arranged in a vector $\mathbf{w}$. The expectation–maximization (EM) algorithm [5, 9] estimates together the model parameters and probabilities of the hidden states. It iteratively increases the log–likelihood function of the observed data given the model structure, $\ln f(\mathbf{w}|\boldsymbol{\theta})$, and consists of two steps at each iteration, denoted the expectation and maximization steps, respectively.

The computation scheme of the EM technique can be summarized as follows.

Begin

1. Initialization:
   Choose an initial model estimate $\boldsymbol{\theta}^0$
   Set the iteration counter $l := 0$

2. Repeat

   (a) Compute the joint pmf for the hidden state variables, $p(\mathbf{S}|\mathbf{w}, \boldsymbol{\theta}^l)$

   (b) Compute the pdf estimates, f

   (c) Set $\boldsymbol{\theta}^{l+1} := \underset{\boldsymbol{\theta}}{\text{argmax}}\ \mathrm{E}[\ln f(\mathbf{w}, \mathbf{S}|\boldsymbol{\theta})|\mathbf{w}, \boldsymbol{\theta}^l]$

   (d) Set $l := l + 1$

   until the convergence toward the maximum likelihood estimate.

End

This algorithm may be simplified by taking into account only the most likely path instead of considering all the possible state transitions.

The computation of the wavelet coefficients can then be done after the training of the HMT model. For a given tree, the minimum mean–square estimate of x is found to be

$$\begin{aligned}\hat{x}(n) &= E[x(n)|\mathbf{y}, \boldsymbol{\theta}] \\ &= \sum_m p(S_i = m|\mathbf{y}, \boldsymbol{\theta}) \frac{\sigma_m^2(n)}{\sigma_m^2(n) + \sigma^2} y(n),\end{aligned} \tag{22}$$

where

$$\sigma_m^2(n) = \begin{cases} \vartheta_m^2(n) - \sigma^2, & \text{if } \vartheta_m^2(n) \geq \sigma^2 \\ 0 & \text{otherwise} \end{cases} \tag{23}$$

and ϑ_m^2 is the variance of the noisy wavelet coefficients in the mth state. At this point, the signal of interest can be restored by performing the inverse transform of the wavelet coefficients. Fig. 4 shows the block diagram of the overall HMT based system, which can be used for the noise reduction.

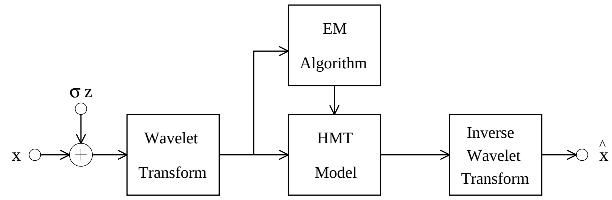

Figure 4: Structure of the HMT based noise reduction.

5. APPLICATIONS

In the case of a length$-N$ signal x in a noise, the $L-$scale wavelet transform of the noisy data results in $N2^{-L}$ trees of wavelet coefficients. The 512×512 Lena picture is adopted for the different applications. The noisy data were generated from this image by adding white Gaussian noise. The mean–square error (MSE) between the clean image, $\mathbf{I}$, and the estimated image, $\hat{\mathbf{I}}$, can be computed as

$$MSE = (\frac{1}{N^2}) \sum_{i=1}^{N} \sum_{j=1}^{N} [\mathbf{I}(i,j) - \hat{\mathbf{I}}(i,j)]^2 \tag{24}$$

and the peak signal–to–noise ratio (PSNR) is

$$\text{PSNR(in dB)} = -10 \log_{10}(\text{MSE}/\max\{\mathbf{I}(i,j)\}^2), \tag{25}$$

where i and j determine the pixel location. For an image with $[0, 255]$ gray–level range, $\max\{\mathbf{I}(i,j)\} = 255$. The following methods were considered in addition to the approach based on the HMT model. The first method is the hard thresholding of wavelet coefficients using a constant threshold for all scales and the second one is the Wiener noise reduction algorithm. Using the HMT technique, the minimum MSE can be improved in the range of $1 - 1.5$dB in comparison to classical denoising algorithms (hard thresholding and Wiener methods). Because the HMT model captures the edge structure of a given image, the edges are retained in the denoised image even for high level noises.

The visual quality of the data estimation achieved by the HMT based algorithm can be appreciated by comparing the images shown in Figs. 5, 6, 7. The computation was done with compactly supported length-8 wavelet functions and the PSNR improvement between the noisy and recovered images is about 9dB.

6. CONCLUSION

The problem of recovering data from its noisy copy is addressed. The use of the HMT structure, which can concisely capture the image characteristics in the

Figure 5: Original image.

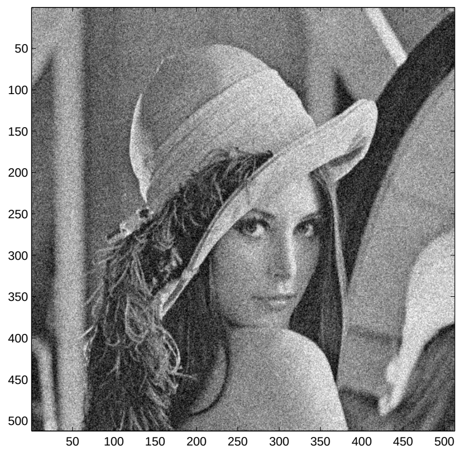

Figure 6: Image corrupted by noise ($PSNR =$ 20.0748dB).

Figure 7: Image recovered after noise reduction ($PSNR = 29.0508$dB).

wavelet domain, results in an efficient noise reduction technique. The EM algorithm was adopted to fit the hidden Markov model to the received signal. A comparison of quantitative and qualitative results for test images demonstrates the improved noise suppression performance with respect to previous image denoising methods. Numerical examples are given to illustrate the performance of the HMT based algorithm.

7. REFERENCES

[1] S. Mallat and W. L. Hwang, “Singularity Detection and Processing with Wavelets”, IEEE Trans. on Info. Theory, Vol. 38, pp. 617–643, Mar. 1992.

[2] D. L. Donoho and I. M. Johnstone, “Ideal Spatial Adaptation by Wavelet Shrinkage”, *Biometrika*, Vol. 81, pp. 425–455, 1994.

[3] D. Donoho, “De–noising by Soft–Thresholding”, IEEE Trans. on Info. Theory, Vol. 41, pp. 613–627, May 1995.

[4] G. P. Nason, “Choice of the Threshold Parameter in Wavelet Function Estimation”, in *Wavelets and Statistics*, A. Antoniadis and G. Oppenheim Eds., Springer Verlag, New York, 1995, pp. 261–280.

[5] O. Ronen, J. R. Rohlicek and M. Ostendorf, “Parameter Estimation of Dependence Tree Models Using the EM Algorithm”, IEEE Signal Processing Lett., Vol. 2, pp. 157–159, Aug. 1995.

[6] C. A. Bouman and M. Shapiro, “A Multiscale Random Field Model for Bayesian Image Segmentation”, IEEE Trans. on Image Processing, Vol. 3, pp. 162–177, Mar. 1994.

[7] M. S. Crouse, R. D. Nowak and R. G. Baraniuk, “Hidden Markov Models for Wavelet–based Signal Processing”, in Proc. 30th Assilomar Conf., Pacific Grove, CA, Nov. 1996.

[8] L. R. Rabiner, “A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition”, Proc. of the IEEE, Vol. 77, pp. 257–286, Feb. 1989.

[9] M.–Y. Zou and R. Unbehauen, “Image Debluring”, in *Diagnosis Optimization Techniques*, C. T. Leondes Eds., Gordon and Breach Science Publishers, Amsterdam, 1997, pp. 147–204.