WAVELET-BASED HIDDEN MARKOV TREES FOR IMAGE NOISE REDUCTION

E. Hošťálková, A. Procházka

Institute of Chemical Technology, Prague
Department of Computing and Control Engineering

Abstract

In the field of signal processing, the Discrete Wavelet Transform (DWT) has proved very useful for recovering signals from additive Gaussian noise by the means of wavelet thresholding. During this procedure, wavelet coefficients with small magnitudes are set to zero, however, usually without taking into account their mutual dependencies. The Hidden Markov Models (HMM) are designed to capture such dependencies by modelling the statistical properties of the coefficients. In this paper, we process a testing intensity image with added Gaussian noise. To compute the hidden Markov models parameters, we employ the iterative expectation-maximization (EM) training algorithm. The outcome of the training process is used for estimation of the noise-free image which is reconstructed from the recalculated wavelet coefficients. The above technique is compared with the NormalShrink method of adaptive threshold computation and outperforms this technique in our experiments.

1 Introduction

The Discrete Wavelet Transform (DWT) is broadly and successfully used for signal estimation by wavelet shrinkage [3]. The shrinkage algorithm consists of wavelet decomposition of the noisy signal observation, thresholding the wavelet coefficients with an estimated threshold value, and subsequent wavelet reconstruction using the altered wavelet coefficients along with the preserved scaling coefficients.

The shrinkage technique may vary according to the thresholding function (hard, soft, or other), the formula for the threshold calculation, and whether it is applied globally for all wavelet coefficients or adaptively using different thresholds for different levels or subbands. In general, shrinkage methods ignore mutual dependencies between DWT coefficients, and thus assume the DWT to de-correlate signals thoroughly. This, however, is not a correct assumption as shown in [2], since the DWT coefficients reveal persistence and clustering [3].

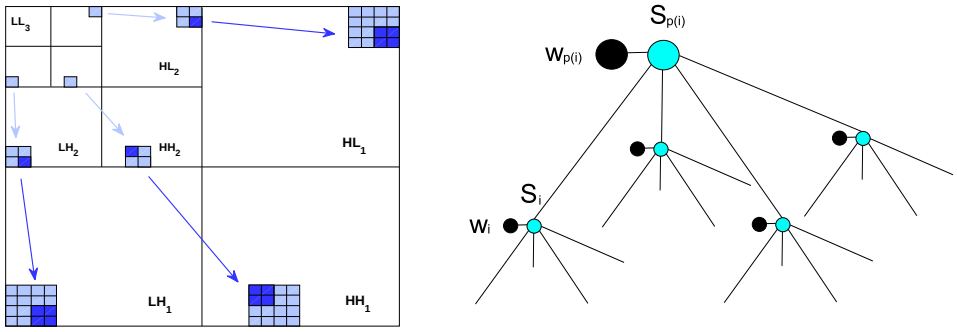

Figure 1: The persistence property of wavelet coefficients. In the 2-dimensional decomposition hierarchy, each parent coefficient $p(i)$ has four children i. The HMT model connects the hidden states S_i and $S_{p(i)}$ rather then the actual coefficients values w_i and $w_{p(i)}$

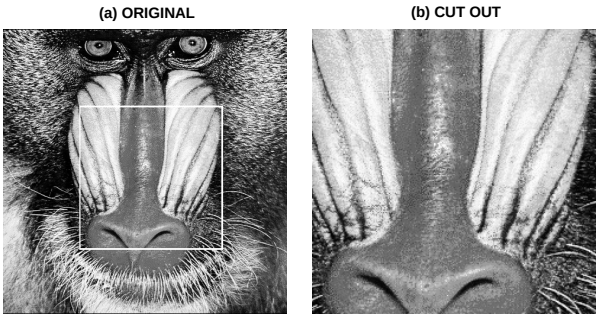

Figure 2: Mandrill image **(a)** and a 240×240 cut out normalized to the intensity range $\langle 0; 1 \rangle$ **(b)**

The *persistence* property denotes strong parent-child relations in the wavelet decomposition hierarchy. The relative size of the coefficients propagates through their children across scale as outlined in Fig. 1. Due to the *clustering* property, we may expect large (or small) coefficients in the neighborhood of a large (or small) coefficient within the same scale.

The latter property is captured by the hidden Markov chains models while ignoring the former. For our purposes, we choose a modelling framework which reflects both these properties - the Hidden Markov Trees (HMT). Apart from noise reduction discussed in this paper, the HMT models are widely used in edge detection, texture recognition, and other applications [2, 1, 6].

1.1 HMT of Wavelet Coefficients

As said above, the HMT models are designed to capture mutual wavelet coefficients dependencies through modelling the statistical properties of the coefficients. Markovian dependencies tie together the hidden states assigned to the coefficients rather than their values, which are thus treated as independent of all variables given the hidden state.

For real images, histograms of the DWT coefficients reveal *sparsity*, which means that the shape of the marginal probability distribution for each wavelet coefficient value is peaky and heavy tailed with relatively few large coefficients corresponding to singularities and many small ones from smooth regions. Hence the marginal distribution of each coefficient node i is modeled as a mixture of Gaussian conditional distributions $G(\mu_{i,m}, \sigma^2_{i,m})$. In many applications, a 2-component mixture proves sufficient.

As displayed in Fig. 3, each of the two conditional distributions (with a smaller variance $\sigma^2_{i,1}$ and a larger variance $\sigma^2_{i,2}$) is associated with one of the two hidden states S taking on values $m = 1, 2$ with the probability mass function (pmf) $p(S_i = m)$. Then, the overall density function

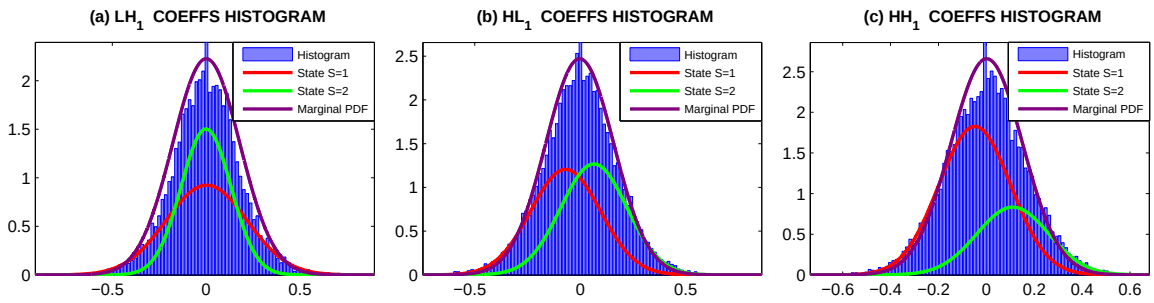

Figure 3: Non-Gaussian marginal densities for all subbands at level 1 obtained via the HMT models. A histogram of the LH coefficients **(a)**, HL coefficients **(b)**, and HH coefficients **(c)** along with the respective conditional densities of the two states (for the noise mean $\mu_n = 0.05$ and variance $\sigma^2_n = 0.03$ in the spatial domain)

is given as

$$f(w_i) = p(S_i = m)\, f(w_i \,|\, S_i = m) \tag{1}$$

where the conditional probability $f(w_i \,|\, S_i = m)$ of the coefficients value w_i given the state S_i corresponds to the Gaussian distribution

$$f(w_i \,|\, S_i = m) = \frac{1}{\sqrt{2\pi\sigma_{i,m}^2}}\; exp\left(-\frac{(w_i - \mu_{i,m})^2}{2\,\sigma_{i,m}^2}\right) \tag{2}$$

For images, each parent coefficient in the HMT hierarchy has four children. Owing to *persistence*, the relative size of the coefficients propagates across scale. To describe these dependencies, the 2-state HMT model uses the state transition probabilities $f(S_i = m \,|\, S_{p(i)} = n))$ between the hidden states S_i of the children given that of the parent $S_{p(i)}$

$$f(S_i = m \,|\, S_{p(i)} = n)) = \begin{pmatrix} f(S_i = 1 \,|\, S_{p(i)} = 1) & f(S_i = 1 \,|\, S_{p(i)} = 2) \\ f(S_i = 2 \,|\, S_{p(i)} = 1) & f(S_i = 2 \,|\, S_{p(i)} = 2) \end{pmatrix} \tag{3}$$

where according to the persistence assumption $f(S_i = 1 \,|\, S_{p(i)} = 1) \gg f(S_i = 2 \,|\, S_{p(i)} = 1)$ and $f(S_i = 2 \,|\, S_{p(i)} = 2) \gg f(S_i = 1 \,|\, S_{p(i)} = 2)$.

In this paper, the DWT wavelet coefficients are modeled using three independent HMT models. In this way, we *tie* together all trees belonging to each of the three detail subbands to decrease the computation complexity and prevent overfitting to the data. The model parameters θ are computed via the iterative expectation-maximization (EM) training algorithm described in detail in [2]. The algorithm consists of two steps. In the E step, the state information propagates upwards and downwards through the tree. In the M step, the model parameters θ are recalculated and then input into next iteration.

1.2 Noise Reduction

In this paper, we deal with denoising of signals containing additive Independent Identically Distributed (iid) Gaussian noise. In the wavelet domain, a noisy wavelet coefficient observation w_i is given by

$$w_i = y_i + n_i \tag{4}$$

where y stands for the desired noise-free signal and n for iid Gaussian noise.

Each of the three HMT models trained in the previous section is exploited for image noise reduction as follows. As derived by the chain rule of conditional expectation, the conditional mean estimate of y_i, given the noise observation w_i and the state s_i [2]

$$E[y_i \,|\, \mathbf{w}, \theta] = \sum_{m=1}^{M} p(S_i = m \,|\, \mathbf{w}, \theta) \cdot \frac{\sigma_{i,m}^2}{\sigma_n^2 + \sigma_{i,m}^2} \cdot w_i \tag{5}$$

The hidden state probabilities $p(S_i \,|\, \mathbf{w}, \theta)$ given the parameters vector θ and observed wavelet coefficients values $\mathbf{w}$ are, same as the variance $\sigma_{i,m}^2$, common to all coefficients in a given subband. As the only unknown remains the noise variance σ_n^2, which can be obtained through the Median Absolute Deviation (MAD) estimator [3]

$$\hat{\sigma_n}^{mad} = \frac{median\{|w_1^{hh1}|, |w_2^{hh1}|, \ldots, |w_{N/4}^{hh1}|\}}{0.6745} \tag{6}$$

where N is the image size and $|w_n^{hh1}|$ is the absolute value of the n-th coefficient of the HH_1 subband, which contains the highest frequencies, and thus is supposed to be noise dominated.

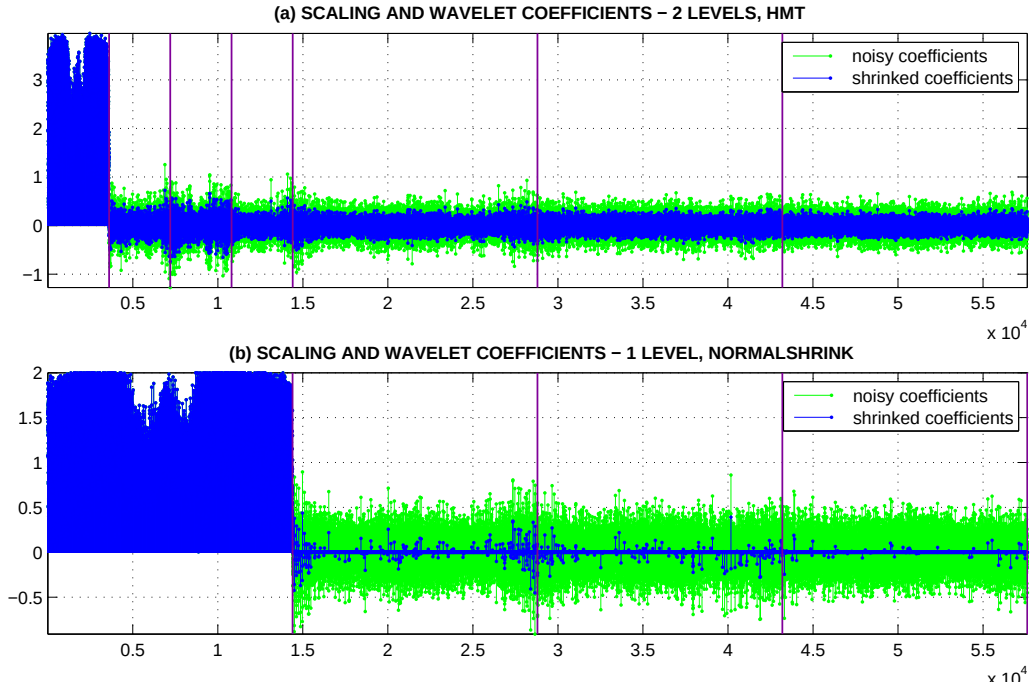

Figure 4: Altering wavelet coefficients by exploiting the HMT model **(a)** and the NormalShrink threshold estimate **(b)**. The Haar DWT coefficients of the noisy image are displayed in green and the altered ones in blue (for the same noisy image as in Fig. 3)

The constant in the denominator applies to iid Gaussian noise. The median approach is robust against large deviations of noise variance.

Now, we are able to compute new values of the wavelet coefficients and use them for DWT reconstruction while keeping the scaling coefficients unchanged as depicted in Fig. 4a.

Fig. 4b displays coefficients processed by the *NormalShrink* method proposed by [4]. This shrinkage technique is subband-adaptive, uses relation (6) for the noise variance estimation, and employs the soft thresholding function. Fig. 2 shows a cut out of the *mandrill* image which we use as testing data.

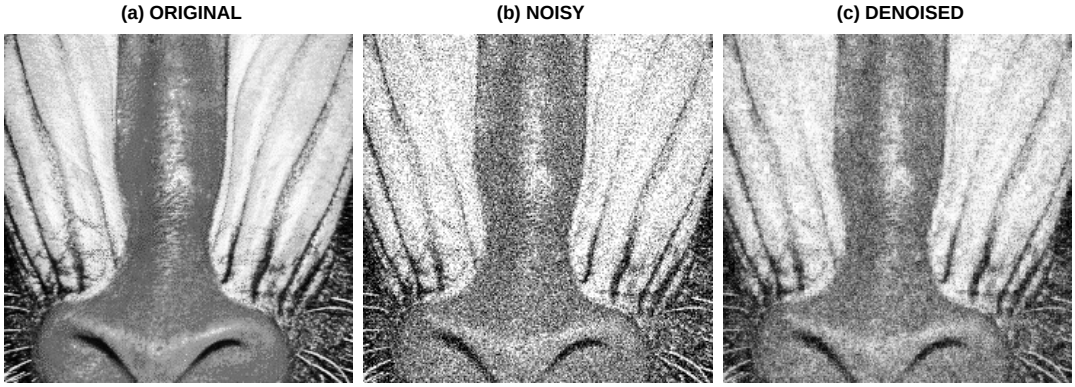

Figure 5: Noise reduction via the HTM models. The original image (the cut out from the mandrill image) **(a)**, the same image with additional iid Gaussian noise ($\mu_n = 0.05$, $\sigma_n^2 = 0.03$) **(b)**, and the result of HMT-based denoising **(c)**

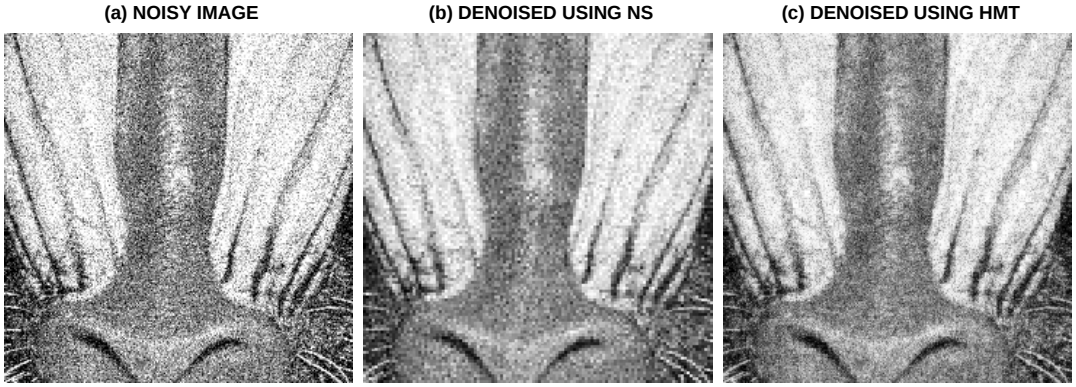

Figure 6: Noise reduction via NormalShrink and the HTM models. The noisy image (the same one as in 5) **(a)**, and the result of NormalShrink **(b)**, and HMT-based denoising **(c)**

1.3 Results

Our experiments, nevertheless limited to only one testing image, verified the expectations derived form literature [2]. The comparison of the HMT-based and the NormalShrink method is summarized in the following table.

Table 1: Residual Images Parameters in Our Noise Reduction Experiments

Noise		*NormalShrink*		*HMT*	
μ_n $[10^{-2}]$	σ_n^2 $[10^{-2}]$	μ $[10^{-2}]$	σ^2 $[10^{-2}]$	μ $[10^{-2}]$	σ^2 $[10^{-2}]$
5.00	3.00	0.04	2.18	1.12	0.60
0.00	1.00	0.00	1.12	0.16	0.32
5.00	1.00	0.46	1.04	1.00	0.32

In case of the HMT-based method, we decomposed the signal to the second level. The NormalShrink technique performed better for single-level decomposition according both to numerical and visual evaluation.

Fig. 6 displays an example of using of the both denoising techniques. We may also visually compare the denoising results in Fig. 7 and conclude, that the HMT-based technique outperforms the other method in preserving image edges.

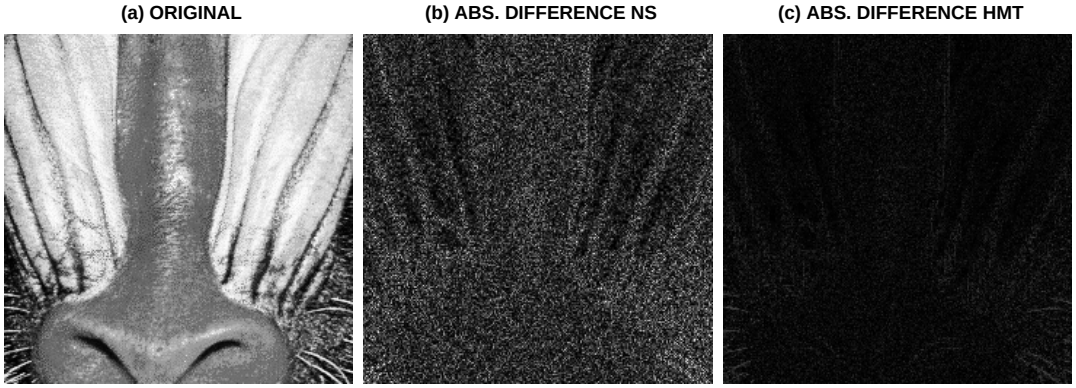

Figure 7: Absolute values difference images for the NormalShrink and the HTM denoising experiments. The original image **(a)**, and the result of NormalShrink method (normalized to the range $\langle 0; 1 \rangle$) **(b)**, and the HMT method (displayed proportionally to the previous image) **(c)**

In our future work, we intend to exploit the HMT models for noise reduction in biomedical images. Instead od the DWT, it will be advantageous to employ the Dual-Tree Complex Wavelet Transform (DTCWT) [6], which is approximately shift invariant and its coefficients magnitudes do not oscillate across scale at the location of a singularity and provides near linear phase encoding.

ACKNOWLEDGEMENTS

The paper has been supported by the Research grant No. MSM 6046137306.

References

[1] H. Choi and R. G. Baraniuk. Multiscale image segmentation using wavelet domain hidden markov models. In *Proceedings of the IEEE International Conf. on Image Processing*, pages 1309 – 1321. IEEE, 2001.

[2] M. S. Crouse, R. D. Nowak, and R. G. Baraniuk. Wavelet-based statistical signal processing using hidden markov models. *IEEE Transactions on Signal Processing*, 46(4):886 – 902, 1998.

[3] D. B. Percival and A. T. Walden. *Wavelet Methods for Time Series Analysis*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, New York, U.S.A., 2006.

[4] L. Kaur, S. Gupta, and R. C. Chauhan. Image denoising using wavelet thresholding. In *Third Conference on Computer Vision, Graphics and Image Processing, India*, pages 1 – 4, 2002.

[5] J. K. Romberg, H. Choi, and R. G. Baraniuk. Multiscale edge grammars for complex wavelet transforms. In *Proceedings of the IEEE International Conf. on Image Processing*, pages 614 – 617. IEEE, 2001.

[6] C. W. Shaffrey, N. G. Kingsbury, and I. H. Jermyn. Unsupervised image segmentation via markov trees and complex wavelets. In *Proceedings of the IEEE International Conf. on Image Processing, Rochester, USA*, pages 801 – 804. IEEE, 2002.