



## Image Denoising through Self-Organizing Feature Map Based on Wavelet-Domain HMMs

Jianxin Dai & Yaqin Jiang

School of Mathematics and Physics

Nanjing University of Posts and Telecommunications

Nanjing 210046, China

E-mail: [daijx\\_js@126.com](mailto:daijx_js@126.com)

### Abstract

Although the Wavelet-domain hidden Markov Models (HMMs) can powerfully preserve the image edge information, it lacks local dependency information. According to the deficiency, a novel image denoising method based HMMs through the self-organizing feature map(SOFM) which exploits spatial local correlation among image neighbouring wavelet coefficients is proposed in this paper. SOFM algorithms is popular for unsupervised learning and data clustering and can capture persistence properties of wavelet coefficients. Experimental results show that the performance of the proposed method is more practicable and more effective to suppress additive white Gaussian noise and preserve the details of the image.

**Keywords:** Self-organizing feature maps, Image denoising, Hidden markov models

### 1. Introduction

Images are easy to be contaminated with noise, either because of the data acquisition process, or because of naturally occurring phenomena. Therefore, image denoising is an important processing task, both as a process itself and as a component in other processes before received images can be used in applications. Denoising images corrupted by additive white Gaussian noise (AWGN) is a classical problem to the image processing community. The aim of image denoising is to remove the noise while keeping the signal features as much as possible.

A lot of work has been done on wavelet for denoising in both the signal processing and statistics communities. However, existing methods implicitly treat each wavelet coefficient as though it is independent of all others. These models are unrealistic for many real world signals.

Recently, the wavelet-domain hidden Markov tree (HMT) model is proposed based on the following three statistical properties of the wavelet coefficients:

(1)Non-Gaussian distribution: The marginal distribution of the magnitude of the complex wavelet coefficients can be well modeled by using a mixture of two-state Rayleigh distributions. The choice for using the Rayleigh mixture model instead of the Gaussian mixture model was based upon the fact that the real and imaginary parts of the complex wavelet coefficients may be slightly correlated, and therefore only the magnitudes of the complex wavelet coefficients will present a nearly shift-invariant property, but not the phase.

(2)Persistency: Large/small wavelet coefficients related to pixels in the image tend to propagate through scales of the quad trees. Therefore, a state variable is defined for each wavelet coefficients that associates the coefficients with one of the two Rayleigh marginal distributions [one with small (S) and the other with large (L) variance]. The HMT model is then constructed by connecting the state variables (L and S) across scales using the Expectation-Maximization (EM) algorithm.

(3)Clustering: Adjacent wavelet coefficients of a particular large/small coefficient are very likely to share the same state (large/small).

The superior results of HMT model denoising have demonstrated that significant performance gains can be achieved by exploiting dependencies among coefficients. The whole parameters of HMT model can be estimated by Expectation Maximization (EM) algorithm. EM algorithm is a kind of Max-Likelihood Estimation (MLE) algorithm whose target is to find a set of hidden states to maximize the probability of observations. If the initial values for parameters of HMM

are known, EM iterates between estimating the probability for the state (Expectation) and updating the model given the state probabilities (Maximization) until the HMM converges. EM algorithm is an efficient algorithm to find local optimum parameters from an initial value. However, it is difficult to gain both reliable and local estimate results of the EM algorithm in signal denoising, because reliable estimate results require that the number of samples are big enough, while local estimate result requires to consider only neighboring wavelet coefficients. The samples are a group of identically independent distribution (iid) data that are used to train the parameters of HMM. Therefore, we should attentively design a model structure to keep a suitable number of samples for training the parameter of our model. Of course, we typically have a single noisy signal observation in signal denoising. Therefore, in order to ensure reliable parameter estimation for the signal we must share statistical information among related wavelet coefficients, and the problem becomes how to share statistical information to gain both reliable and local estimate results of EM. To accomplish this problem, the standard assumption of wavelet-domain HMMs is that all wavelet coefficients and state variables within a common scale are iid. In this assumption, wavelet coefficients within a common scale can be considered as samples of the parameter estimate, then the parameter of a HMM can be trained by all wavelet coefficients within a common scale. Thus the standard assumption provides a powerful tool to ensure reliable parameter estimation for the signal, but it lacks local dependency information. The reason is parameters of HMMs within a common scale are the same that does not consider the positions of the samples. To strike a balance between enough related wavelet coefficients and spatial adaptability, we must carefully and attentively design the related relationship among wavelet coefficients.

This paper proposes a novel structure of HMMs through SOFM (SOFM-HMM) which exploits spatial local related statistical information among image neighbouring wavelet coefficients to improve the spatial adaptability of HMMs. SOFM is one of the major unsupervised artificial neural network models and often used to learn certain useful features found in their learning process. Thus ,the new framework provides a flexible conditional probability model for efficient learning and expressing the dependencies in wavelet transforms.

The remaining sections are organized as follows. Section 2 briefly reviews wavelet-domain HMMs. Section 3 studies the SOFM algorithm. Section 4 explains the proposed image denoising method and it is followed by experimental results comparing with various wavelet denoising methods in section 5 and conclusion in section 6, respectively.

**2. The Wavelet Domain Hidden Markov Models Theory**

The wavelet domain HMMs theory was first proposed by Crouse based on the statistical properties introduced in Section 1. In practice, the distribution of the wavelet coefficients is described as non-Gaussian with a peak centered at 0 and a heavy tail. In this paper, Gaussian mixture models are adapted to approximate this non-Gaussian density property. In this model, each wavelet coefficient  $w$  is associate with a discrete value named hidden state,  $S = 0, 1, \dots, M - 1$ , and the coefficient's probability density function pdf  $f_{w|S}(w|S = m)$  given its hidden state can be described as a Gaussian distribution with mean  $\mu_m$  and variance  $\sigma_m^2$ . Using these assumptions, the distribution of the coefficients can be described as  $f_w(w) = \sum_{m=0}^{M-1} p_s(m) f_{w|S}(w|S = m)$ . In hidden markov model, the coefficients are assumed to be independent

with each other given the hidden states. Because the wavelet transform have a natural quardtree structure, so, by using the across scale persistence property, the HMT established a Markov chain structure which can avoid the two dimensional lattice's uncausal feature, and the computation consumption of the parameters estimation can be efficiently reduced. The HMT is shown in figure1. From it, we can see the dependencies across scale. Based on the M state Gaussian mixture model, the parameter set  $\theta_{HMT}$  of the model is gives as  $\{p_{S_i}(m), \epsilon_{i,\rho(i)}^{mn}, \mu_{im}, \sigma_{im}^2 | m = 0, 1, \dots, M - 1\}$  Where  $\epsilon_{i,\rho(i)}^{mn}$  denote the transition probability of node  $\rho(i)$  in state  $n$  to its son node  $i$  in state  $m$  and  $p_{S_i}(m)$  denote the initial distribution of node  $i$  in state  $m$ . In practice, we use a two-states, zero-mean Gaussian mixture models and the low-variance Gaussian model is used to model the low-energy wavelet coefficients, the high-variance Gaussian to high-energy coefficients.

The whole parameters of HMT can be estimated by EM algorithm. In this algorithm, the following conditional likelihoods are defined:

$$\beta_i(m) = f(T_i | S_i = m, \theta). \tag{1}$$

$$\beta_{i,\rho(i)}(m) = f(T_i | S_{\rho(i)} = m, \theta) \tag{2}$$

$$\beta_{\rho(i)i}(m) = f(T_{\rho(i)i} | S_{\rho(i)} = m, \theta). \tag{3}$$

And the joint probability function is defined:

$$\alpha_i(m) = p(S_i = m, T_{1i} | \theta). \tag{4}$$

Then in the E step, by upward-downward scanning, (1)-(4) leads to the desired conditional probabilities

$$p(S_i = m|W, \theta) = \frac{\alpha_i(m)\beta_i(m)}{\sum_{n=1}^M \alpha_i(n)\beta_i(n)} \tag{5}$$

$$p(S_i = m, S_{\rho(i)} = n|W, \theta) = \frac{\alpha_{\rho(i)}(m)\beta_{\rho(i)i}(m)\beta_i(m)\varepsilon_{i,\rho(i)}^{nm}}{\sum_{n=1}^M \alpha_i(n)\beta_i(n)} \tag{6}$$

To maximize the parameter set  $\theta$ , we perform the within trees tying to finish the M step and get the maximized parameter set

$$ps_i(m) = \frac{1}{K} \sum_{k=1}^K p(S_i^k = m|W^k, \theta) \tag{7}$$

$$\varepsilon_{i,\rho(i)}^{mn} = \frac{1}{K p_{S_{\rho(i)}}(m)} \sum_{k=1}^K p(S_i^k = n, S_{\rho(i)}^k = m|W^k, \theta) \tag{8}$$

$$\sigma_{i,m}^2 = \frac{1}{K p_{S_i}(m)} \sum_{k=1}^K (w_i^k - \mu_{i,m})^2 p(S_i^k = m|W^k, \theta) \tag{9}$$

Where  $K$  denote the numbers of the coefficients of one subband in one scale, so the M step is finished by assuming that these coefficients are in same distribution and they are independent of each other. The following section will describe our proposed model in detail.

### 3. SOFM Algorithm

The self-organizing feature map (SOFM) is one of the major unsupervised artificial neural network models and often used to learn certain useful features found in their learning process. It basically provides a way for cluster analysis by producing a mapping of high dimensional input vectors onto a two-dimensional output space while preserving topological relations as faithfully as possible. After appropriate training iterations, the similar input items are grouped spatially close to one another. As such, the resulting map is capable of performing the clustering task in a completely unsupervised fashion.

SOFM is a topology preserving nonlinear transformation. Each neuron,  $j, 1 \leq j \leq l$  is connected to the input through a synaptic weight vector  $M_j = [m_{j1}, \dots, m_{jm}]^T$ . At each iteration, SOFM finds the best-matching (winning) neuron  $j$  by minimizing the following cost function:  $V(X) = \arg \min_{j=1}^{j=l} \|X_j - M_j\|$ ,

Where  $X_j$  belongs to an  $m$ -dimensional input space,  $\|\cdot\|$  denotes the Euclidean distance, while the update of the synaptic weight vectors follows:

$$M_j(t+1) = M_j(t) + \alpha(t) N_j(t) [X_j(t) - M_j(t)], \quad j = 1, \dots, l. \tag{10}$$

where  $\alpha(t), 0 < \alpha(t) < 1$ , designates the learning rate factor;  $N_j(t)$  designates the decreasing neighbourhood function centred on the winner and means the number of neuron at the  $t$  learning times in the competitive layer. Although the algorithm is simple, its convergence and accuracy depend on the selection of the neighbourhood function, the topology of the output space, a scheme for decreasing the learning rate parameter, and the total number of neuronal units.

The SOFM algorithm can be summarized in the following basic steps.

- 1) Randomly select a training vector  $X_j$  from corpus. Make sure the initial value  $\alpha(0)$ , make sure the initial value  $N_j(0)$ , make sure the total learning times  $T$  ..
- 2) Find the winning neuron  $j$  with the weight  $M_j$  which is closest to  $X_j$  by minimizing the cost function  $V(X)$ .
- 3) Adjust the weight between the competitive layer neuron and input vector which in the neighborhood  $N_j(t)$  of the competitive layer by formula (10). In practice, in order to better reflect the relationship between the input patterns and the neural weight vector in competition and improve the classification accuracy, we change the weight adjusting formula as:

$$M_j(t+1) = M_j(t) + \alpha(t) N_j(t) \exp\{-\rho_{xm}\} [X_j(t) - M_j(t)], \quad j = 1, \dots, l. \tag{11}$$

Where  $\rho_{xm}$  is correlation coefficient and reflects the correlation of  $X_j$  and  $M_j$ , it changes between  $[-1, 1]$ , and  $\rho_{xm}$  is designated as:

$$\rho_{XM} = \frac{E\left\{\left[X_j - E(X_j)\right]\left[M_j - E(M_j)\right]\right\}}{\sqrt{D(X_j)}\sqrt{D(M_j)}} \tag{12}$$

Where  $E(X_j) = \frac{1}{n} \sum x_{ji}$ ,  $E(M_j) = \frac{1}{n} \sum m_{ji}$ ,

$$D(X_j) = E(X_j^2) - [E(X_j)]^2, \quad D(M_j) = E(M_j^2) - [E(M_j)]^2$$

4) Renew  $\alpha(t)$  and  $N_j(t)$

$$\alpha(t) = \alpha(0) \left(1 - \frac{t}{T}\right) \tag{13}$$

$$N_j(t) = INT \left\lfloor N_j(0) \left(1 - \frac{t}{T}\right) \right\rfloor \tag{14}$$

5)  $t = t + 1$  and return step (2), until  $t = T$ .

**4. Image Denoising based SOFM-HMM**

After doing a 2D wavelet transform, we get four frequency subbands, namely, LL, LH, HL, and HH at every decomposition level. Because most energy of noise concentrates in high frequency, only wavelet coefficients in the high frequency levels need to be thresholded. Since the correlation among image wavelet coefficients is stronger and the correlation among white-noising wavelet coefficients is much weaker, wavelet coefficients can be easily classified through application of the SOFM algorithm. Wavelet coefficients in every high frequency level (LH, HL, and HH) are trained respectively, such resolves the problem of shortage of memory, and at the same time improves the training speed. SOFM-HMM has smaller number of observations than standard HMM, such as HMT, because the observations are separated into several classes using SOFM, and the number of observations in each class is smaller than the number of observations in all the classes. For this reason, SOFM-HMM has lower computation complexity than HMT. The other factor that affects the convergent speed is the structure of HMM. As the dependency model underlying the HMMs becomes more complicated, each iteration of the EM algorithm becomes slower, and the algorithm may take longer to converge. Hence, it is important to keep the HMMs as simple as possible. Thus, SOFM-HMM yields a flexible framework for signal denoising that strikes a balance between enough samples and spatial adaptability of the training algorithm. Once a trained HMM is obtained, estimation of the true signal wavelet coefficients (denoising) is straightforward. We can get the modified wavelet coefficients using the estimate formula. The final signal estimate (denoised signal) is computed as the inverse wavelet transform of these estimates of the signal wavelet coefficients.

Briefly, the SOFM-HMM algorithm of image denoising can be described as follows:

- 1) Perform forward 2D wavelet decomposition on the noisy image.
- 2) Apply the SOFM-HMM algorithm to train wavelet coefficients in every high frequency level (LH, HL, and HH) respectively.
- 3) Apply EM algorithm to estimate the whole parameters of SOFM-HMM.
- 4) Compute the estimation of the wavelet coefficients.
- 5) Perform inverse 2D wavelet transform of the estimation of the wavelet coefficients.

From our experiments we find that SOFM-HMM performs better than VisuShrink and HMM.

**5. Experimental Results**

The 512 x 512 standard grayscale image Lena is used in experiment. For comparison, we implement SOFM-HMM, VisuShrink and HMT. The Daubechies wavelet with 8 vanishing moments is used for the wavelet decomposition. For comparison, we implement VisuShrink, HMT, SOFM-HMM. VisuShrink is the universal soft-thresholding denoising technique. For different Gaussian white noise levels, the experiential results in Peak Signal to Noise Ratio (PSNR) are shown in Table 1 for denoising images Lena. The PSNR is defined as

$$PSNR = -10 \log_{10} \frac{\sum_{i,j} (B(i,j) - A(i,j))^2}{n^2 \max_{i,j} A(i,j)^2} \tag{15}$$

where  $B(i,j)$  is the denoised image and  $A(i,j)$  is the noise-free image. The first column in this table is the PSNR of the original noisy images, while other columns are the PSNR of the denoised images by using different denoising methods. From Table 1 we can see that SOFM-HMM outperforms VisuShrink and HMT for all cases.

By studying the denoised images in Figure 2, we see that SOFM-HMM produces smoother and clearer denoised images.

**6. Conclusions and Future Work**

In this paper, we propose a wavelet image denoising method based HMMs through the SOFM algorithm which exploits spatial local correlation among image neighbouring wavelet coefficients. The main contribution lies in the classification strategy based on the SOFM algorithm. Experimental results have show that SOFM-HMM gives better results when the noise variance gets larger. However, the main drawbacks of this approach is the high computational cost, which grows exponentially with the dimensions of the output space. Future work may be done by extending this idea to the multiwavelet case.

**References**

Achim A, Kuruoğlu E E, (2005). Image denoising using bivariate  $\alpha$ -stable distributions in the complex wavelet domain [J]. *IEEE Signal Processing Letters*. 12(1): 17-20.

Crouse M.S., Nowak R.D.and Baraniuk R.G., (1998). Wavelet-based statistical signal processing using hidden Markov models. *IEEE Trans. Signal Processing*. 46 (4), pp.886-902,.

D.L.Donoho, (1995).Denoising by soft-thresholding. *IEEE Trans.on Information Theory*. vol. 41, pp. 613–627.

Donoho D L, Johnstone I M, (1994). Ideal spatial adaptation via wavelet shrinkage [J]. *Biometrika*. 81(3): 425-455.

Ferrari Ricardo J, Winsor R ,(2005). Digital radiographic image denoising via wavelet-ased hidden Markov model estimation [J]. *Journal of Digital Imaging*. 18(2): 154-167.

Romberg J K, Choi H, Baraniuk R G, (2001). Bayesian tree-structured image modeling using wavelet-domain hidden Markov models [J]. *IEEE Transactions on Image Processing*. 10(7): 1056-1068.

T. T. Cai , B. W. Silverman, (2001). Incorporating information on neighbouring coefficients into wavelet estimation. *Sankhya: The Indian Journal of Statistics*. Vol. 63, Series B, Pt. 2, pp. 127-148.

Table 1. The PSNR (dB) of the noisy images of Lena and the denoised images with different denoising methods.

Noisy Image	VisuShrink	HMT	SOFM-HMM
25.28	25.19	27.12	27.65
21.20	24.03	25.38	25.93
17.32	22.31	24.89	25.34
15.12	21.16	22.09	23.83

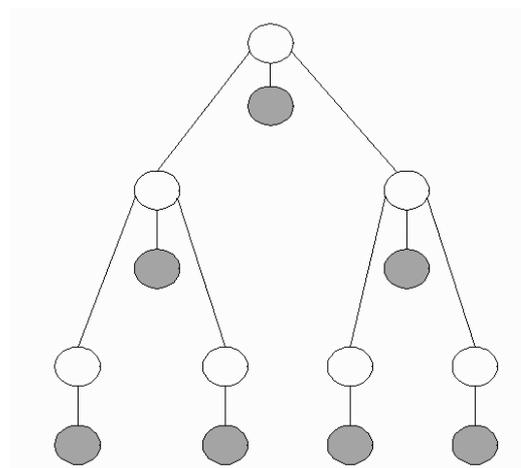


Figure 1. HMT where white nodes represent hidden states and dark nodes represent wavelet coefficients.

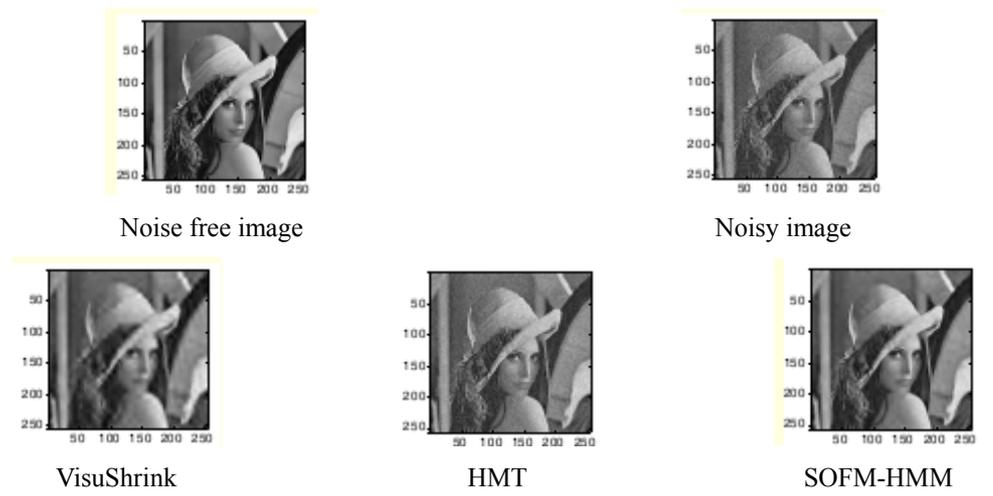


Figure 2. Image denoising by using different methods on an noisy image with PSNR = 21dB.