JOURNAL OF INFORMATION SCIENCE AND ENGINEERING 21, 1261-1275 (2005)

Short Paper

Contextual Hidden Markov Tree Model for Signal Denoising

MING-YU SHIH AND DIN-CHANG TSENG
Institute of Computer Science and Information Engineering
National Central University
Chungli, 320 Taiwan
E-mail: tsengdc@ip.csie.ncu.edu.tw

The hidden Markov tree (*HMT*) model is a novel statistical model for signal and image processing in the wavelet domain. The *HMT* model captures the interscale persistence property of wavelet coefficients, but includes only a tiny intrascale clustering property of wavelet coefficients. In this paper, we propose the contextual hidden Markov tree (*CHMT*) model to enhance the clustering property of the *HMT* model by adding extended coefficients associated with the wavelet coefficients. The extended coefficients are regarded as leaves to link wavelet coefficients in the wavelet tree model without destroying the wavelet persistence property; hence, the training approach of the *HMT* model can be modified to estimate the parameters of the *CHMT* model. In experiments, the proposed *CHMT* model produced better results than the *HMT* model for signal denoising. Furthermore, the *CHMT* model needs fewer iterations of training than the *HMT* model to get the same denoised results.

Keywords: contextual analysis, hidden Markov model, hidden Markov tree model, signal denoising, wavelet transform

1. INTRODUCTION

Recently, the study of signal and image processing using statistical models is growing. Rabiner [13] gave a tutorial on hidden Markov models, and Willsky [18] proposed a review about multiresolution Markov models for signal and image processing. The wavelet transform records differences of neighboring signals for several different scales [14]; it has the locality, multiresolution, compression, clustering, and persistence properties [6] and is well suited for signal and image analysis. Wavelet-based statistical models regard wavelet coefficients as a random realization. Most statistical models use a Gaussian statistical model for analysis, but almost all wavelet coefficients of signals do not have really a Gaussian distribution [6]; they are compressed and typically described by a peak at zero and heavy-tailed non-Gaussian distribution. Thus, Gaussian mixture distribution models were employed to approximate the non-Gaussian distribution [3].

Received July 18, 2003; revised July 23, 2004; accepted January 24, 2005.
Communicated by C. C. Jay Kuo.

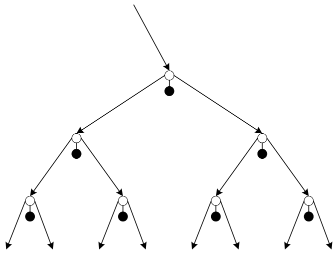
In the Gaussian mixture models (*GMM*), a wavelet coefficient is associated with a hidden state variable to describe various distribution states, where each state follows a Gaussian distribution. The simplest approach is the independent mixture (*IM*) model [3, 7, 12]. This model treats each state variable as an independent random variable, and there are no relations across scales and within a scale. To exploit the wavelet properties, hidden Markov models (*HMM*'s) were proposed to capture intrascale [11, 13], interscale [1, 4, 6, 8, 14], or both [2, 5, 9, 10, 17] dependencies. These techniques model signals more accurately and produce better results than the simple *IM* model produces.

Crouse *et al.* [6] proposed the hidden Markov tree (*HMT*) model in the wavelet domain. Based on *GMM*, the *HMT* model captures the interscale dependencies using the Markovian relation between the hidden states of the parent-child wavelet coefficients in the tree structure as shown in Fig. 1 (a); the model has been successfully applied in signal and image denoising [6, 8, 14] and image segmentation [1, 4, 8]. Although the *HMT* model fully covers the interscale persistence property, which includes only tiny intrascale clustering property among coefficients under the common ancestor coefficient, the relationship of adjacent coefficients with different parents needs to be enhanced.

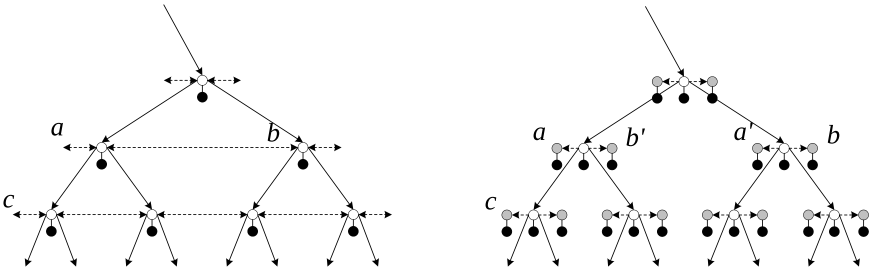
Wavelet coefficients have the persistence and clustering properties, i.e.; a large or small wavelet coefficient tends to have large or small values in its neighbors and across scales. Hence, simultaneously capturing inter- and intrascale dependencies can model wavelet coefficients more accurately than only capturing inter- or intrascale dependency can. Crouse and Baraniuk [5] proposed an efficient contextual hidden Markov model (*CHMM*) to capture both inter- and intrascale dependencies based on the context vectors between wavelet coefficients. They defined the mixing probability for a wavelet coefficient in the *GMM* based on the context vector which is composed of adjacent coefficients within a scale and across scales. The model only considered local relations; thus dependencies exist only between a coefficient and the connected coefficients with the context vectors and do not propagate to other coefficients. The denoising results of the *CHMM* are generally worse than that of the *HMT* model presented in their experiments [5]. Fan and Xia [9] utilized other local characteristic to improve the *CHMM*, but dependencies are not propagated in the whole wavelet hierarchy.

To capture both inter- and intrascale dependencies in *GMM*, the hidden Markov chain model can be used [13]. One simple way is to connect the hidden states of a coefficient with those of neighboring coefficients within and across scales as the one 1-D structure in Fig. 1 (b) shows. Although this model can capture inter- and intrascale dependencies, the prior probability of one coefficient would be heavily affected by other non-inherited coefficients. For example in Fig. 1 (b), nodes *a* and *c* will be intensely affected by node *b* and its ancestor and descendant nodes in the iterative parameter training, and the dependencies violate the wavelet hierarchy characteristics in which the persistence property is stronger than the clustering property in a wavelet hierarchy [17]. Hence, the model is not suitable for modeling wavelet coefficients. In above description, a coefficient along with its state variables is called a node.

In this paper, we propose a contextual hidden Markov tree (*CHMT*) model by adding intrascale dependencies in the *HMT* model to capture the strong wavelet persistence property and the weaker clustering property for signal denoising. Instead of directly connecting hidden states of two neighboring coefficients in a scale, we add extended coefficients associated with wavelet coefficients; then the hidden states of an extended



(a) The standard *HMT* model.



(b) Intrascale dependencies are added between pairs of neighboring hidden state variables.

(c) The proposed approach adds intrascale dependencies from real state variables to extended state variables (gray dots).

Fig. 1. Interscale dependencies between two adjacent scales and intrascale dependencies within a scale are shown as solid and dash lines, respectively, where wavelet coefficients (black dots) are associated with hidden state variables (white dots).

coefficient are connected with those of the neighboring wavelet coefficients as shown in Fig. 1 (c). An extended coefficient has the same value as its original coefficient, but has its own state variable and parameters such as means, variances, and transition functions which are independent of those of the original wavelet coefficient.

In Fig. 1 (c), nodes a' and b' are the extended nodes of nodes a and b, respectively. Nodes a and b' affect each other, and nodes a' and b affect each other. Node a' has the same value as node a, and b' has the same value as node b. Nodes a and b affect each other, but nodes a and c are not affected by the ancestor/descendant nodes of node b in the iterative parameter training. Consequentially, the weak intrascale dependencies are included while the strong interscale dependencies are kept. In other words, we maintain the benefit of the wavelet persistence property and add intrascale dependencies to respond to the cluster property. Using extended coefficients improves the likelihood estimation of parameters for all coefficients, which more completely describes the dependencies among all coefficients in the wavelet hierarchy. The proposed *CHMT* model is an extension of the *HMT* model. The extended coefficients do not destroy the wavelet persistence property. Moreover, the training algorithm can be easily modified from that of the *HMT* model to estimate our parameters.

The remaining sections of this paper are organized as follows. The standard *HMT* model is introduced in section 2. The proposed *CHMT* structure, *EM* training algorithm, and denoising method are presented in section 3. Experiments are reported in section 4. Conclusions are given in section 5.

2. STANDARD *HMT* MODEL

The *HMT* model as shown in Fig. 1 (a) uses a Gaussian mixture distribution with hidden state variables to model non-Gaussian wavelet coefficients. We regard wavelet coefficient w_i as an observation of random variable W. A discrete hidden state variable S_i is associated with wavelet coefficient w_i. The pair (w_i, S_i) is called a node in the tree model.

A Gaussian probability density function (*pdf*) of random variable W is

$$g(W;\ \mu,\ \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{(W-\mu)^2}{2\sigma^2}\right\}, \tag{1}$$

where μ and σ are the mean and standard derivation of W, respectively. Wavelet coefficient w_i is associated with M-state Gaussian mixture conditional *pdf* values

$$f_{W|S_i}(w_i \mid S_i = m) = g(w_i;\ \mu_{i,m},\ \sigma_{i,m}^2),\ \ m = 1,\ 2,\ \ldots,\ M. \tag{2}$$

Assume that S_i has a probability mass function (*pmf*) $p_{S_i}(m)$, then the overall *pdf* $f_W(w_i)$ is determined by

$$f_W(w_i) = \sum_{m=1}^{M} p_{S_i}(m)\, f_{W|S_i}(w_i \mid S_i = m)\,. \tag{3}$$

Assume that the wavelet decomposition has L level and the coarsest-scaled level is 1. Let $w_{p(i)}$ be the parent coefficient of coefficient w_i and be associated with the hidden state variable $S_{p(i)}$. Combining with the M-state Gaussian mixture model, the parameter set θ_{HMT} of the *HMT* model is defined as

$$\theta_{HMT} = \left\{p_{S_1}(m),\ \varepsilon_{i,p(i)}^{mn},\ \mu_{i,m},\ \sigma_{i,m}^2 \mid m,\ n = 1,\ ...,\ M\right\}, \tag{4}$$

where

i. $p_{S_1}(m)$ is the *pmf* value of the root state variable S_1 in state m;

ii. $\varepsilon_{i,p(i)}^{mn} = p_{S_i|S_{p(i)}}(m \mid S_{p(i)} = n)$ is the conditional probability that S_i is in state m given that $S_{p(i)}$ is in state n;

iii. $\mu_{i,m}$ and $\sigma_{i,m}^2$ are the mean and variance of the Gaussian mixture model of the wavelet coefficient w_i given S_i is in state m.

The parameter set θ_{HMT} can be determined by the *EM* algorithm [6]. In practice, a two-state, zero-mean Gaussian mixture model is sufficient for modeling the wavelet coefficients of most real-world images [6]. The low-variance Gaussian model is used to model the wavelet coefficients having high energy, and the high-variance Gaussian is used to model the low-energy coefficients.

3. CONTEXTUAL HIDDEN MARKOV TREE MODEL

The contextual *HMT* (*CHMT*) model is modified from the *HMT* model by adding dependencies between coefficients in a scale. In the model, each wavelet coefficient is associated with extended coefficients that are copied from the adjacent coefficients and are treated as tree leaves in each scale as shown in Fig. 1 (c). More specifically, if the coefficient set is $\{w_1, w_2, \ldots, w_k\}$ in a wavelet subband, we add coefficients for every coefficient to form the extended coefficient set $\{w_1, w_{e(1,2)}, w_{e(2,1)}, w_2, w_{e(2,2)}, \ldots, w_{e(i,1)}, w_i, w_{e(i,2)}, \ldots, w_{e(k,1)}, w_k\}$, where $w_{e(1,2)} = w_2$, $w_{e(2,1)} = w_1$, $w_{e(2,2)} = w_3$, …, $w_{e(i,1)} = w_{i-1}$, $w_{e(i,2)} = w_{i+1}$, …, and $w_{e(k,1)} = w_{k-1}$. Each original or extended coefficient is regarded as an observation of random variable W, and each coefficient is associated with M-state Gaussian mixture conditional *pdf*. Extended coefficient $w_{e(i,j)}$ is associated with a hidden state variable $S_{e(i,j)}$, where $j = 1, 2$. For convenience in describing a tree structure, we define node $e(i, j)$ to be the pair $(w_{e(i,j)}, S_{e(i,j)})$. We add Markovian dependencies between hidden state variable S_i and each extended hidden state variable $S_{e(i,j)}$; and each intrascale dependency is represented by a transition probability δ.

To match the description of extended coefficients, we use child nodes instead of parent nodes in the proposed *CHMT* model. In the *CHMT* model, the interscale dependencies are described similarly to those of the *HMT* model. The extended coefficients are treated as the tree leaves and have no child coefficients associated with them. Let i be an original node, $c(i, j)$ be the jth child node of node i. In a wavelet subband, one node has two children. We use the transition probability ε to represent the interscale dependencies between hidden state variables S_i and $S_{c(i,j)}$. If the wavelet tree has N coefficients and each coefficient is modeled by the M-state Gaussian mixture model, the parameter set θ_{CHMT} of the *CHMT* model is

$$\theta_{CHMT} = \left\{ p_{S_1}(m),\ \varepsilon^{nm}_{c(i,j),i},\ \delta^{nm}_{e(i,j),i},\ \mu_{i,m},\ \sigma^2_{i,m} \mid i = 1, \ldots, N;\ m,\ n = 1, \ldots, M \right\}, \tag{5}$$

where

i. $p_{S_1}(m)$ is the probability of the root state variable S_1 in state m;

ii. $\varepsilon^{nm}_{c(i,j),i} = p_{S_{c(i,j)}|S_i}(n \mid S_i = m)$ is the (conditional) transition probability of node i to its child node $c(i, j)$ when $S_{c(i,j)}$ is in state n given S_i is in state m;

iii. $\delta^{nm}_{e(i,j),i} = p_{S_{e(i,j)}|S_i}(n \mid S_i = m)$ is the (conditional) transition probability of node i to its extended node $e(i, j)$ when $S_{e(i,j)}$ is in state n given S_i is in state m;

iv. $\mu_{i,m}$ and $\sigma^2_{i,m}$ are the mean and variance of the Gaussian model of wavelet coefficient w_i given S_i in state m for all original and extended coefficients.

In the proposed *CHMT* model, the transition probabilities ε's are used to define the interscale conditional dependencies, and δ's for defining the intrascale conditional dependencies. The extended hidden state variable $S_{e(i,j)}$ is conditionally independent of the entire structure given S_i; that is, the intrascale conditional dependency only influences the adjacent coefficients and is not propagated to other coefficients. Hence, we prepare the merits of the *HMT* model and improve the dependency description for wavelet coefficients.

3.1 Parameter Training using EM Algorithm

The expectation maximization (*EM*) algorithm modified from the algorithm of Crouse *et al.* [6] is used to iteratively adjust parameters based on the maximum likelihood (*ML*) principle. The algorithm consists of three steps: up expectation, down expectation, and maximization steps. Based on the current model parameters θ_{CHMT}, the up step determines the likelihood of subtrees from the finest scale to the coarsest scale and the down step determines the joint probability of subtrees from the coarsest scale to the finest scale. Combining these two steps, all state transition probabilities are evaluated. Finally, we get the new model parameters $\hat{\theta}_{CHMT}$ by calculating the mean of the expectation results in the maximization step.

3.1.1 Definition

Let T_i be the tree with root at wavelet coefficient w_i and $T_{i/j}$ be the tree T_i deprived of its subtree T_j. Let T_1 be the entire tree. The conditional likelihoods β_i and the joint probability functions α_i for node i are then

$$\beta_i(m) = f(T_i \mid S_i = m,\ \theta_{CHMT}), \tag{6}$$

$$\beta_{i/c(i,j)}(m) = f(T_{i/c(i,j)} \mid S_i = m,\ \theta_{CHMT}), \tag{7}$$

$$\beta_{i/e(i,j)}(m) = f(T_{i/e(i,j)} \mid S_i = m,\ \theta_{CHMT}), \tag{8}$$

and

$$\alpha_i(m) = p(S_i = m,\ T_{1/i} \mid \theta_{CHMT}); \tag{9}$$

hence, the joint probability of node i in state m and entire tree can be calculated by

$$p(S_i = m,\ T_1 \mid \theta_{CHMT}) = \alpha_i(m)\beta_i(m), \tag{10}$$

and the entire likelihood of all observed wavelet coefficients is

$$f(T_1 \mid \theta_{CHMT}) = \sum_{m=1}^{M} p(S_i = m,\ T_1 \mid \theta_{CHMT}) = \sum_{m=1}^{M} \alpha_i(m)\beta_i(m) \tag{11}$$

for any node i. After α's and β's are determined, the *pmf*s for each node and each chain of two nodes can then be calculated as:

$$p(S_i = m \mid \boldsymbol{w},\ \theta_{CHMT}) = \frac{\alpha_i(m)\beta_i(m)}{\sum_{n=1}^{M} \alpha_i(n)\beta_i(n)}, \tag{12}$$

$$p(S_{c(i,j)} = n,\ S_i = m \mid \boldsymbol{w},\ \theta_{CHMT}) = \frac{\beta_{c(i,j)}(n)\varepsilon_{c(i,j),i}^{nm}\alpha_i(m)\beta_{i/c(i,j)}(m)}{\sum_{n=1}^{M}\sum_{m=1}^{M} \beta_{c(i,j)}(n)\varepsilon_{c(i,j),i}^{nm}\alpha_i(m)\beta_{i/c(i,j)}(m)}, \tag{13}$$

and

$$p(S_{e(i,j)} = n,\ S_i = m \mid \boldsymbol{w},\ \theta_{CHMT}) = \frac{\beta_{e(i,j)}(n)\delta^{nm}_{e(i.j),i}\alpha_i(m)\beta_{i/e(i,j)}(m)}{\sum_{n=1}^{M}\sum_{m=1}^{M}\beta_{e(i,j)}(n)\delta^{nm}_{e(i,j),i}\alpha_i(m)\beta_{i/e(i,j)}(m)}, \tag{14}$$

where $\boldsymbol{w}$ is the set of all wavelet coefficients in the tree.

3.1.2 The up step

Based on the Gaussian distribution, the conditional likelihoods β for all extended nodes in all states are

$$\beta_{e(i,j)}(n) = g(w_{e(i,j)};\ \mu_{e(i,j),n},\ \sigma^2_{e(i,j),n}),\ \ n = 1,\ \ldots,\ M, \tag{15}$$

the β's of all child nodes of a node i at the finest scale are

$$\beta_{c(i,j)}(n) = 1,\ \ n = 1,\ 2,\ \ldots,\ M, \tag{16}$$

and the β's of all other nodes i are

$$\beta_i(m) = g(w_i;\ \mu_{i,m},\ \sigma^2_{i,m})\prod_{j=1}^{nc}\left(\sum_{n=1}^{M}\varepsilon^{nm}_{c(i,j),i}\beta_{c(i,j)}(n)\right)\prod_{j=1}^{ne}\left(\sum_{n=1}^{M}\delta^{nm}_{e(i,j),i}\beta_{e(i,j)}(n)\right), \tag{17}$$

where nc and ne are the numbers of child nodes and extended nodes of node i. It should be noted that the number of child node is fixed for each tree node, but the number extended nodes varies.

3.1.3 The down step

Initially, the joint probability α of each node at the coarsest scale is set to

$$\alpha_1(m) = p_{S_1}(m),\ \ m = 1,\ \ldots,\ M. \tag{18}$$

Then α's of all child nodes and extended nodes of node i are computed by

$$\alpha_{c(i,j)}(n) = \sum_{m=1}^{M}\left(\varepsilon^{nm}_{c(i,j),i}\alpha_i(m)\beta_{i/c(i,j)}(m)\right) \tag{19}$$

with

$$\beta_{i/c(i,j)}(m) = \frac{\beta_i(m)}{\sum_{k=1}^{M}\varepsilon^{km}_{c(i,j),i}\beta_{c(i,j)}(k)} \tag{20}$$

and

$$\alpha_{e(i,j)}(n) = \sum_{m=1}^{M} \left(\delta^{nm}_{e(i,j),i} \alpha_i(m) \beta_{i/e(i,j)}(m) \right) \tag{21}$$

with

$$\beta_{i/e(i,j)}(m) = \frac{\beta_i(m)}{\sum_{k=1}^{M} \delta^{km}_{e(i,j),i} \beta_{e(i,j)}(k)}. \tag{22}$$

3.1.4 The maximization step

Suppose that there are K trees in a wavelet hierarchy and all trees have the same characteristics. The maximization step estimates new parameter set $\hat{\theta}_{CHMT}$ for the next iteration by averaging the parameters associated with the K trees,

$$\hat{p}_{S_1}(m) = \frac{1}{K} \sum_{k=1}^{K} p(S_1^k = m \mid \boldsymbol{w}^k,\ \theta_{CHMT}), \tag{23}$$

$$\hat{\varepsilon}^{nm}_{c(i,j),i} = \frac{\sum_{k=1}^{K} p(S^k_{c(i,j)} = n,\ S_i^k = m \mid \boldsymbol{w}^k,\ \theta_{CHMT})}{\sum_{n=1}^{M} \sum_{k=1}^{K} p(S^k_{c(i,j)} = n,\ S_i^k = m \mid \boldsymbol{w}^k,\ \theta_{CHMT})}, \tag{24}$$

$$\hat{\delta}^{nm}_{e(i,j),i} = \frac{\sum_{k=1}^{K} p(S^k_{e(i,j)} = n,\ S_i^k = m \mid \boldsymbol{w}^k,\ \theta_{CHMT})}{\sum_{n=1}^{M} \sum_{k=1}^{K} p(S^k_{e(i,j)} = n,\ S_i^k = m \mid \boldsymbol{w}^k,\ \theta_{CHMT})}, \tag{25}$$

$$\hat{\mu}_{i,m} = \frac{\sum_{k=1}^{K} w_i^k\, p(S_i^k = m \mid \boldsymbol{w}^k,\ \theta_{CHMT})}{\sum_{k=1}^{K} p(S_i^k = m \mid \boldsymbol{w}^k,\ \theta_{CHMT})}, \tag{26}$$

and

$$\hat{\sigma}^2_{i,m} = \frac{\sum_{k=1}^{K} (w_i^k - \mu_{i,m})^2\, p(S_i^k = m \mid \boldsymbol{w}^k,\ \theta_{CHMT})}{\sum_{k=1}^{K} p(S_i^k = m \mid \boldsymbol{w}^k,\ \theta_{CHMT})}. \tag{27}$$

3.1.5 Universal tying

To better estimate the correct parameters, we need a large amount of training data to prevent overfitting problem [6]. However, there are only a few observations with which to estimate the parameters for a coefficient. Thus we use the universal tying strategy [14] to link parameters for estimation. In the wavelet structure, universal tying means that the similar parameters of all coefficients in a subband are tied as a parameter because the coefficients in a subband have the same property. This strategy reduces the number of parameters; in other words, and, hence, effectively increases the amount of training data for relative to the number of parameters that need to be estimated.

Based on the universal tying, the tying parameter set of the *CHMT* model is given as

$$\theta_{CHMT}^{T}=\left\{p_{S_1}(m),\ \varepsilon_{c(l),l}^{nm},\ \delta_{e(l),l}^{nm},\ \mu_{l,m},\ \sigma_{l,m}^{2}\mid l=1,\ \ldots,L;\ m,\ n=1,\ \ldots,\ M\right\}, \tag{28}$$

where L is the number of scales of all wavelet trees, $c(l)$ denotes the related child nodes, and $e(l)$ denotes the related extended nodes. In the up and down steps, all conditional likelihoods and joint probabilities of each node i are calculated along with associated parameters in the scale where node i is located. In the maximization step, the new parameters at scale l are acquired from posterior probabilities of all nodes at scale l. For example, the new transition probability $\hat{\varepsilon}_{c(l),l}^{nm}$ is estimated as

$$\hat{\varepsilon}_{c(l),l}^{nm}=\frac{\sum_{k=1}^{K}\sum_{i\in\text{scale }l}\sum_{j\in c(l)}p(S_{c(i,j)}^{k}=n,\ S_{i}^{k}=m\mid \boldsymbol{w}^{k},\ \theta_{CHMT}^{T})}{\sum_{n=1}^{M}\sum_{k=1}^{K}\sum_{i\in\text{scale }l}\sum_{j\in c(l)}p(S_{c(i,j)}^{k}=n,\ S_{i}^{k}=m\mid \boldsymbol{w}^{k},\ \theta_{CHMT}^{T})}. \tag{29}$$

The other parameters can be estimated in the same way.

3.2 Denoising

We use the proposed *CHMT* model and the empirical Bayesian Estimation [6] for signal denoising. Suppose that white Gaussian noise is added to signals. Based on the wavelet transform of signals, the noise problem can be described as

$$w_i^k=z_i^k+n_i^k, \tag{30}$$

where w_i^k, z_i^k, and n_i^k are the wavelet coefficients of the noisy signal, original signal, and Gaussian noise of variance σ_n in the k-th tree, respectively.

After training the *CHMT* model, we get the model parameter set θ_{CHMT}. Then we use the maximization step of the *EM* algorithm to calculate the hidden state probabilities $p(S_i^k\mid \boldsymbol{w}^k,\theta_{CHMT})$ for each wavelet coefficient w_i^k of the noisy signals. Suppose that the variance of the noisy signals in the mth state is $\gamma_{i,m}$, then the conditional estimate of the mean for z_i^k [6] is

$$E\left[z_i^k \mid \mathbf{w}^k,\ \theta_{CHMT}\right] = \sum_{m=1}^{M} p(S_i^k = m \mid \mathbf{w}^k,\ \theta_{CHMT})\ \frac{(\gamma_{i,m}^2 - \sigma_n^2)}{\gamma_{i,m}^2} + w_i^k, \tag{31}$$

where $(x)_+ = \begin{cases} x & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}.$

In the above computation, we assume the noise variance σ_n^2 is known. Actually, we don't know the noise variance in practical applications; however, we can estimate the noise variance from the signals using the median estimate [6] which is performed on the wavelet coefficients of the finest scale as

$$\hat{\sigma}_n^2 = \text{Median}(|w_i|) / 0.6745, \tag{32}$$

where w_i's are wavelet coefficients in the subband HH_1 [15].

4. EXPERIMENTS

We here compare the denoised results of the proposed *CHMT* and standard *HMT* [6] models. Both models use the *EM* algorithm to iteratively train the model parameters, but the proposed *CHMT* model use both inter- and intrascale parameters to describe wavelet coefficients while the *HMT* model use only interscale parameters.

The Daubechies length-8 orthonormal wavelet filters [6, 14] were used to decompose signals, and a two-state zero-mean Gaussian mixture function was used to model the wavelet coefficients of all Markov models. All signals were wavelet decomposed up to the coarsest levels. To reduce the complication of so many parameters, the universal tying strategy [14] was used to tie coefficients in a given scale. Assuming a noisy signal was decomposed into *L* levels, then the universal parameters for each wavelet tree in the *HMT* and *CHMT* models are

$$\theta_{HMT} = \left\{p_{S_1}(m),\ \varepsilon_{c(l),l}^{mn},\ \sigma_{l,m}^2 \mid l = 1,\ ...,\ L;\ m,\ n = 1,\ 2\right\} \tag{33}$$

and

$$\theta_{CHMT} = \left\{p_{S_1}(m),\ \varepsilon_{l,p(l)}^{nm},\ \delta_{e(l),l}^{nm},\ \sigma_{l,m}^2 \mid l = 1,\ ...,\ L;\ m,\ n = 1,\ 2\right\}, \tag{34}$$

respectively, where l denotes a wavelet scale level, m and n are states; and only two states for high and low variance Gaussian functions are considered. The initial values of parameters $p_{S_1}(m)$, $\varepsilon_{c(l),l}^{nm}$, $\varepsilon_{l,p(l)}^{nm}$, and $\delta_{e(l),l}^{nm}$ are set 0.5, all $\sigma_{l,1}^2$ are set to half the variance of wavelet coefficients in scale l, and the other variances $\sigma_{l,2}^2$ are set to twice the variance of wavelet coefficients in scale l.

Four signals, Bumps, Blocks, Doppler, and Heavisine were used for the experiments. Every signal has 1024 samples and was decomposed into 10 levels of wavelet transform. The noisy signals were generated by adding white Gaussian noise of variances 1 and 2.25. The mean square error (*MSE*) between a denoised signal and the original signal was used to evaluate the denoising performance. The denoised results are given in Tables 1 and 2.

Table 1. Comparison of signal denoising results generated by *HMT* and *CHMT* models in *MSE* with 1000 iterations while signals were corrupted by white Gaussian noise of $\sigma_n^2 = 1$.

Signal	Noisy signal	*HMT*	*CHMT*
Bumps	0.9849	0.2714	0.2401
Blocks	1.0220	0.3466	0.3091
Doppler	1.0040	0.1793	0.1594
Heavisine	1.0799	0.1171	0.0666

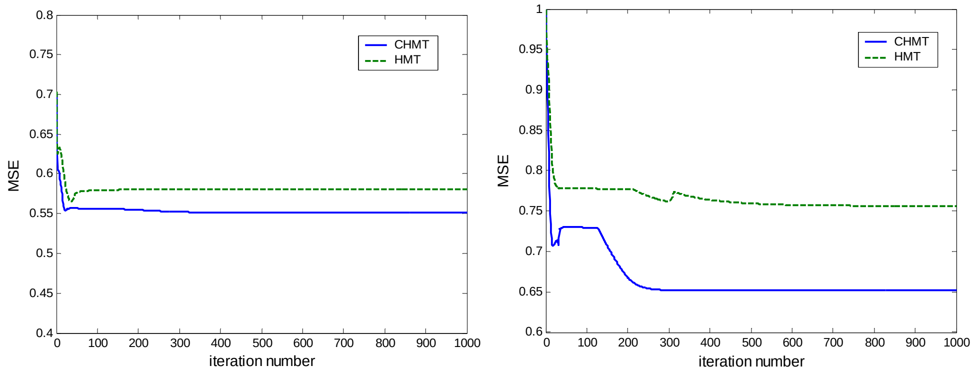
Table 2. Comparison of signal denoiseing results generated by *HMT* and *CHMT* models in *MSE* with 1000 iterations while signals were corrupted by white Gaussian noise of $\sigma_n^2 = 2.25$.

Signal	Noisy signal	*HMT*	*CHMT*
Bumps	2.2612	0.5803	0.5513
Blocks	2.3878	0.7564	0.6519
Doppler	2.2843	0.3612	0.3419
Heavisine	2.3679	0.2838	0.2494

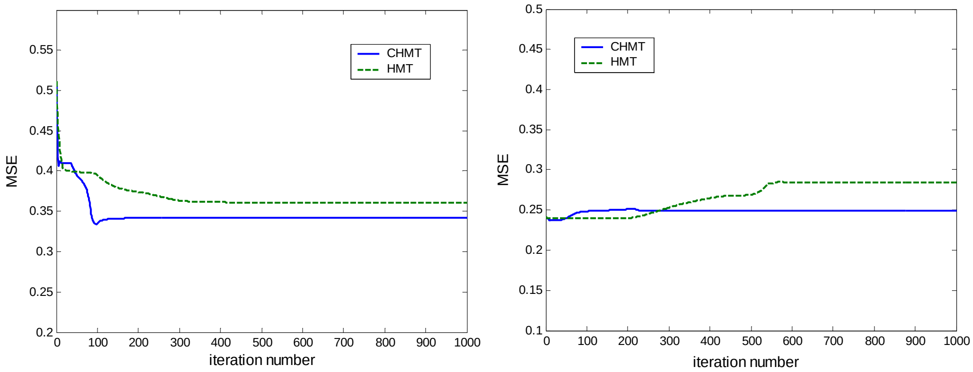
Model parameters and the parameter estimation methods are the two most important factors in the performance of the statistical Markov models. In the parameter estimation, 1000 iterations were used to train parameters in both the *CHMT* and *HMT* models. From the results of denoising shown in Tables 1 and 2, we see that the parameters of the proposed *CHMT* model better take into account the relation between the wavelet coefficients, and thus the *CHMT* model obtained better results.

Next, we compare the convergence speed of iterative parameter estimation for the *HMT* and *CHMT* models as shown in the charts of Fig. 2. In each chart, the horizontal axis indicates the number of iterations and the vertical axis indicates the *MSE* values between the denoised and the original signals. For the first few iterations, the results of both models depend on the initial parameters. However, both models quickly approach better results even if the initial parameters do not fit the signals. The proposed *CHMT* model converges to the expected results more quickly than the *HMT* model; Moreover, even after many iterations, the *HMT* model dose not achieve the results as good as that of the proposed *CHMT* model for all four standard signals.

The *HMT* model produced an unexpected result for the Heavisine signal as shown in Fig. 2 (d). We can use Fig. 3 to explain the phenomenon. There are several high deviation samples in the noisy Heavisine as shown in Fig. 3 (b). The *HMT* model did not make use of intrascale dependency to compress the high level noises as shown in Fig. 3 (c). The proposed *CHMT* model used both inter- and intrascale dependencies for denoising, and the high deviation noises were reduced by the intrascale dependency in the coarse wavelet scales as shown in Fig. 3 (d).

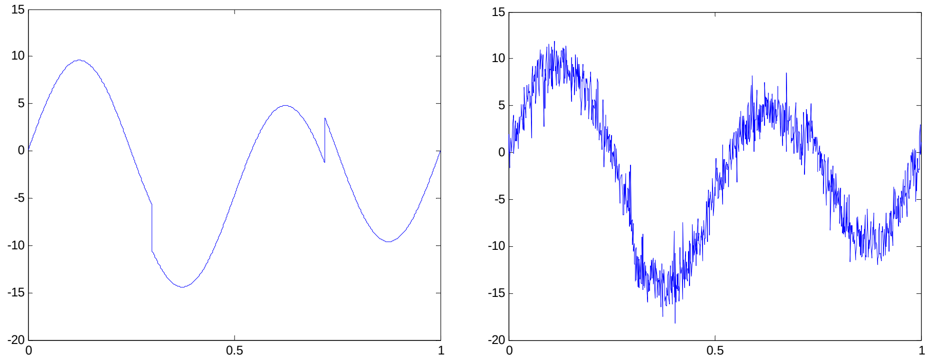


(a) Bumps. (b) Blocks.



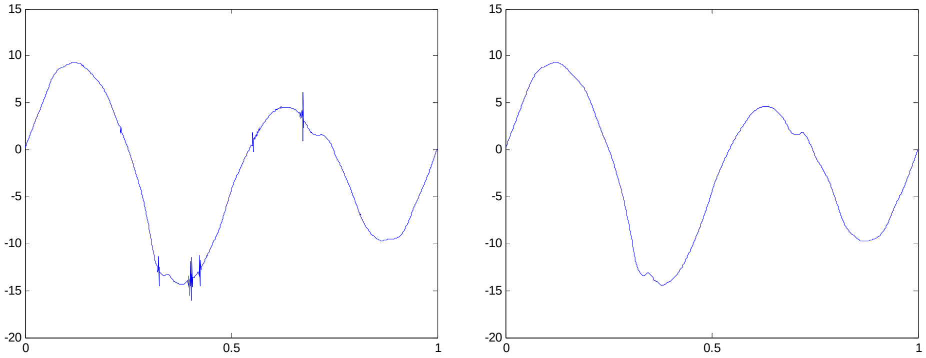
(c) Doppler. (d) Heavisine.

Fig. 2. Comparison of denoising for *HMT* and *CHMT* models with different iterations. All noisy signals were corrupted by white Gaussian noise of $\sigma_n^2 = 2.25$.



(a) The original signal. (b) Noisy signal.

Fig. 3. Heavisine signal.



(c) The denoised signal of the *HMT* model.

(d) The denoised signal of the proposed *CHMT* model.

Fig. 3. (Cont'd) Heavisine signal.

5. CONCLUSIONS

In this paper, we proposed a statistical Markov model called contextual hidden Markov tree (*CHMT*) to model wavelet coefficients by enhancing the clustering property of the *HMT* model [6], and applied the model to signal denoising. In the wavelet domain, we add extended coefficients to establish the intrascale Markovian dependencies to slightly include the wavelet clustering property and maintain the tree dependencies of the *HMT* model to keep the wavelet persistence property. The *EM* algorithm [6] was also modified to train our model parameters and used to estimate the state probabilities for signal denoising.

The experimental results show that the proposed model produces a better description of wavelet coefficients and gets better denoised results than the *HMT* model gets in all processed cases. Furthermore, the proposed *CHMT* model converges to the expected results more quickly than the *HMT* model.

We discussed the proposed *CHMT* model applied to 1-D signals. The *CHMT* model can be extended to 2-D images. The image processing results may be discussed in another article later. From the viewpoint of wavelet hierarchies, 1-D signals and 2-D images have both the interscale and intrascale dependencies. 2-D images have more contextual information, and thus the proposed *CHMT* model may have more significant results.

Wavelet-based hidden Markov model is a solid tool for signal/image processing. In addition to the signal denoising, other applications in segmentation, compression, discrimination, and detection of signals and images will be taken into account in our future work.

REFERENCES

1. C. Bouman and M. Shapiro, “A multiscale random field model for Bayesian image segmentation,” *IEEE Transactions on Image Processing*, Vol. 3, 1994, pp. 162-177.

2. S. G. Chang, B. Yu, and M. Vetterli, "Spatially adaptive wavelet thresholding with context modeling for image denoising," *IEEE Transactions on Image Processing*, Vol. 9, 2000, pp. 1522-1531.
3. H. A. Chipman, E. D. Kolaczyk, and R. E. McCulloch, "Adaptive Bayesian wavelet shrinkage," *Journal of the American Statistical Association*, Vol. 440, 1997, pp. 1413-1421.
4. H. Choi and R. G. Baraniuk, "Multiscale image segmentation using wavelet-domain hidden Markov models," *IEEE Transactions on Image Processing*, Vol. 10, 2001, pp. 1309-1321.
5. M. S. Crouse and R. G. Baraniuk, "Contextual hidden Markov models for wavelet-domain signal processing," in *Proceedings of the 31st Asilomar Conference Signals, System, and Computers*, 1997, pp. 95-100.
6. M. S. Crouse, R. D. Nowak, and R. G. Baraniuk, "Wavelet-based statistical signal processing using hidden Markov models," *IEEE Transactions on Signal Processing*, Vol. 46, 1998, pp. 886-902.
7. D. Donoho and I. Johnstone, "Ideal spatial adaptation via wavelet shrinkage," *Biometrika*, Vol. 81, 1994, pp. 425-455.
8. G. Fan and X. G. Xia, "Wavelet-based statistical image processing using hidden Markov tree model," in *Proceedings of the 2000 Conference on Information Sciences and Systems*, 2000, pp. TA5-31-TA5-36.
9. G. Fan and X. G. Xia, "Image denoising using local contextual hidden Markov model in the wavelet domain," *IEEE Signal Processing Letters*, Vol. 8, 2001, pp. 125-128.
10. J. Liu and P. Moulin, "Image denoising based on scale-space mixture modeling of wavelet coefficients," in *Proceedings of the IEEE International Conference on Image Processing*, 1999, pp. 386-390.
11. M. K. Mihcak, I. Kozintsev, and K. Ramchandran, "Low-complexity image denoising based on statistical modeling of wavelet coefficients," *IEEE Signal Processing Letters*, Vol. 6, 1999, pp. 300-303.
12. P. Moulin and J. Liu, "Analysis of multiresolution image denoising schemes using generalized-Gaussian and complexity priors," *IEEE Transactions on Information Theory*, Vol. 45, 1999, pp. 909-919.
13. L. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," in *Proceedings of the IEEE*, Vol. 77, 1989, pp. 257-285.
14. J. K. Romberg, H. Choi, and R. G. Baraniuk, "Bayesian tree-structured image modeling using wavelet-domain hidden Markov models," *IEEE Transactions on Image Processing*, Vol. 10, 2001, pp. 1056-1068.
15. L. Sendur and I. W. Selesnick, "Bivariate shrinkage with local variance estimation," *IEEE Signal Processing Letters*, Vol. 9, 2002, pp. 438-441.
16. J. M. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients," *IEEE Transactions on Signal Processing*, Vol. 10, 1993, pp. 3445-3462.
17. E. P. Simoncelli, "Statistical models for images: compression, restoration and synthesis," in *Proceedings of the 31st Asilomar Conference Signals, System, and Computers*, 1997, pp. 673-678.
18. A. S. Willsky, "Multiresolution Markov models for signal and image processing," in *Proceedings of the IEEE*, Vol. 90, 2002, pp. 1396-1458.

Ming-Yu Shih (石明于) received his M.S. degree in Computer Science and Information Engineering from National Central University, Chungli, Taiwan. He is currently a Ph.D. student in the same institute. His currently research interests include digital image processing, computer vision, and 3D model reconstruction.

Din-Chang Tseng (曾定章) received his Ph.D. degree in Information Engineering from National Chiao Tung University, Hsinchu, Taiwan, in June 1988. He is currently a professor in the Department of Computer Science and Information Engineering at National Central University, Chungli, Taiwan. He is a member of the IEEE, the Chinese Image Processing and Pattern Recognition Society, and Chinese Geographic Information Association. His current research interests include image processing, computer vision, and virtual reality; especially in the topics: vision-based safety vehicle system, visual synthetic and natural hybrid coding, view-dependent multiresolution terrain modeling, and virtual city.