

# MULTISCALE TEXTURE SEGMENTATION USING HYBRID CONTEXTUAL LABELING TREE

*Guoliang Fan and Xiang-Gen Xia*

Department of Electrical and Computer Engineering  
University of Delaware, Newark, DE 19716, USA  
Email: {glfan,xxia}@ece.udel.edu

## ABSTRACT

Wavelet-domain hidden Markov tree (HMT) model has been recently proposed and applied to image processing, e.g., image segmentation. A new multiscale image segmentation method, called HMTseg, was proposed by Choi and Baraniuk using the wavelet-domain HMT. In this paper, we study the HMTseg algorithm and investigate the Contextual Labeling Tree which is used for the context-based Bayesian interscale fusion of the multiscale classification information. In order to attain more accurate multiscale characterizations with improved segmentation results, we develop three new context structures which have different advantages on the interscale fusion. Then we propose a hybrid context-based interscale fusion algorithm where the three contexts are serially cascaded so that the Bayesian estimation is conducted based on the three contexts respectively and sequentially. The proposed method outperforms the original HMTseg algorithm by improving the accuracies of both texture classification and boundary localization.

## 1. INTRODUCTION

The task of image segmentation aims at separating an image into simple regions with homogeneous behaviors. The homogeneous statistical properties are referred to as “texture”. Recently, many Bayesian estimation techniques have been proposed for modeling the shape of segmented regions and the behavior of pixels in each homogeneous region. In order to achieve reliable segmentation results and accurate boundary localization for a given image, both the large-scale and small-scale behaviors should be investigated and incorporated appropriately. Thus, a natural way to approach this problem is the multiscale analysis [1].

In [2], a new statistical model in the wavelet-domain, hidden Markov tree (HMT), is proposed, which effectively captures the interscale dependencies of the wavelet coefficients across the scales. The HMT model was applied to image segmentation in [3], where a new multiscale image segmentation algorithm, referred to as HMTseg, was proposed. The HMTseg is implemented by three steps, the HMT model training, the maximum likelihood (ML) classification and the context-based interscale fusion. The first step achieves the statistical characterizations of textures. The second step creates the initial ML segmentation results in different scales. The last step computes the maximum a posteriori (MAP)

estimate of the class labels of image pixels via a Contextual Labeling Tree (CLT). In the HMTseg algorithm, the CLT plays an important role for the interscale fusion and is specified by a context structure to characterize the dependencies between the multiscale classification information.

In this paper, we investigate the CLT by considering its effects on the multiscale information characterization. In order to achieve higher accuracies of both texture classification and boundary localization, we develop three new context structures which apply to homogeneous regions or/and texture boundaries, respectively. Based on the three contexts, we propose a Hybrid Contextual Labeling Tree (HCLT) for the Bayesian interscale fusion where the three context structures are cascaded in a serial manner. The new HCLT model is trained by the Expectation Maximization (EM) algorithm three times with respect to the three contexts, respectively and sequentially, so that the MAP estimates of the class labels can be refined step-by-step. The newly proposed algorithm outperforms the original HMTseg algorithm in [3] by improving the accuracies of both texture classification and boundary localization.

## 2. HYBRID CONTEXTUAL LABELING TREE

At first, we briefly review the framework of the HMTseg algorithm, and we refer the reader to [3] for more details.

### 2.1. HMTseg Algorithm

#### 2.1.1. Wavelet-Domain HMT

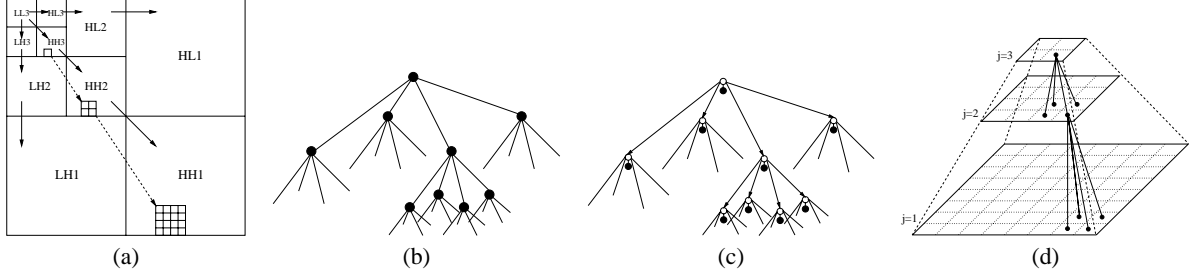
The discrete wavelet transform (DWT) represents a signal or image with both the spatial and frequency characterizations. We illustrate the subbands of the 2-D 3-scale DWT of an image in Fig. 1(a). We can see the quad-tree structure of the wavelet coefficient subtree in each subband  $B$  with  $B = LH, HL, HH$ , as shown in Fig. 1(b), where we can find a nesting structure of the wavelet subtrees. In [2], a new statistical model, called wavelet-domain hidden Markov tree (HMT) model, has been proposed to characterize the joint pdf of wavelet coefficients, as shown in Fig. 1(c). The HMT model is a multidimensional Gaussian mixture model which applies Markovian dependencies between the hidden states along the quad-tree structure. Using a vector  $\theta$  to denote the HMT model parameters, the joint pdf of wavelet coefficients  $\mathbf{w}$  is approximated by  $f(\mathbf{w}|\theta)$  [2].

#### 2.1.2. Multiscale Texture Segmentation

In [3], an image of  $N \times N$  can be recursively divided into four sub-images of same size  $J$  times and represented in a so called

---

This work was partially supported by the 1998 Office of Naval Research (ONR) Young Investigator Program (YIP) under Grant N00014-98-1-0644 and the Air Force Office of Scientific Research (AFOSR) under Grant No. F49620-00-1-0086.



**Fig. 1.** (a) 2-D 3-scale DWT. (b) The quad-tree structure of wavelet subtree rooted in one wavelet coefficient  $w$ . (c) 2-D HMT model where the white node represents the hidden mixture state variable  $S$ . The black node denotes a continuous wavelet coefficient  $W$  which is a random variable. (d) The pyramid representation of the dyadic image blocks.

pyramid representation of  $J$  scales, as shown in Fig. 1(d) which has the same nesting quad-tree structure as the wavelet subtree in Fig. 1(b). Since we desire each dyadic block be associated with a set of specific wavelet coefficient subtrees, the Haar wavelet transform was adopted in [3]. We denote a dyadic block as  $D_{j,i,k}$  which means the  $(i, k)$ th block in the  $j$ th scale of the pyramid representation,  $i, k = 0, \dots, N_j - 1$  with  $N_j = N/2^j$  and  $j = 1, \dots, J$ . From Fig. 1(d), we see the parent block of  $D_{j,i,k}$  is  $D_{j+1, \lfloor i/2 \rfloor, \lfloor k/2 \rfloor}$  where the operation  $\lfloor x \rfloor$  takes the integer part of  $x$ .

Given the Haar DWT  $\mathbf{w}$  of  $N \times N$ , and a set of HMT model parameter  $\theta = \{\theta^{LH}, \theta^{HL}, \theta^{HH}\}$  which consists of three sets of the HMT parameter for the three subbands, respectively, The dyadic block  $D_{j,i,k}$  is associated with three wavelet coefficient subtrees in each subband as,

$$\{\mathcal{T}_{j,i,k}^{LH}, \mathcal{T}_{j,i,k}^{HL}, \mathcal{T}_{j,i,k}^{HH}\},$$

and each subtree is rooted in  $w_{j,i,k}^{LH}$ ,  $w_{j,i,k}^{HL}$  and  $w_{j,i,k}^{HH}$ , respectively.  $w_{j,i,k}^B$  denotes the  $(i, k)$ th wavelet coefficient in the  $j$ th scale of subband  $B$ . Thus the computation of  $f(D_{j,i,k}|\theta)$  is a realization of the HMT model  $\theta$  and is obtained by

$$f(D_{j,i,k}|\theta) = f(\mathcal{T}_{j,i,k}^{LH}|\theta^{LH})f(\mathcal{T}_{j,i,k}^{HL}|\theta^{HL})f(\mathcal{T}_{j,i,k}^{HH}|\theta^{HH}). \quad (1)$$

The computation algorithm of (1) can be found in [2].

### 2.1.3. Implementation

The HMTseg algorithm proposed in [3] was implemented by three steps. Firstly, the HMT model training is performed for the Haar DWT of each texture image and the HMT model parameters are stored in a dictionary. Then, we compute the Haar DWT of the image to be segmented and obtain the pyramid representation. According to (1), we compute the likelihoods of all dyadic blocks with respect to different HMT model parameters and obtain the initial ML multiscale classification results. Finally, the Bayesian interscale fusions are conducted via the CLT using the EM algorithm, and the image segmentation is implemented by computing the MAP estimate of the class label for each dyadic block. The pixel-level segmentation can be obtained by characterizing the pdf of the pixel intensity for each texture image and appending pixels to the pyramid representation as the smallest dyadic blocks.

The Bayesian interscale fusion in the CLT which is specified by a context structure. In [3], the context is used to simplify the determination of the MAP classification of dyadic blocks of the image. The class label of  $D_{j,i,k}$  is denoted as  $C_{j,i,k} = c$ ,  $c = 1, \dots, N_c$  and  $N_c$  is the class number. Each  $D_{j,i,k}$  is associated with a vector context  $V_{j,i,k} = v$ ,  $v = 1, \dots, N_v$  and  $N_v$  is

the number of different context values. Thus the MAP estimate of  $C_{j,i,k}$  is reduced to

$$\hat{C}_{j,i,k} = \arg \max_{c \in \{1, \dots, N_c\}} f(D_{j,i,k}|C_{j,i,k} = c)p_{C_j|V_j}(c|v), \quad (2)$$

where  $p_{C_j|V_j}(c|v)$  is the conditional pmf in the  $j$ th scale, and  $f(D_{j,i,k}|C_{j,i,k} = c)$  is given by (1). In [3], the CLT is specified two parameters, the pmf of the class labels  $p_{C_j}(c)$  and the conditional pmf  $p_{C_j|V_j}(c|v)$ , which can be grouped into a parameter vector  $\alpha$  as

$$\alpha = \{p_{C_j}(c), p_{V_j|C_j}(v|c) | j = 1, \dots, J; c = 1, \dots, N_c; v = 1, \dots, N_v\}. \quad (3)$$

We define  $\mathbf{D}$ ,  $\mathbf{C}$  and  $\mathbf{V}$  as the collections of all  $D_{j,i,k}$ ,  $C_{j,i,k}$  and  $V_{j,i,k}$  in the CLT specified by (3). The problem of interscale fusion can be interpreted as, given  $\mathbf{D}$  and the structure of  $\mathbf{V}$ , we want to find  $\alpha$  and  $\mathbf{C}$  which satisfy (2) and the solution of this problem can be effectively approached by the EM training algorithm described in [3].

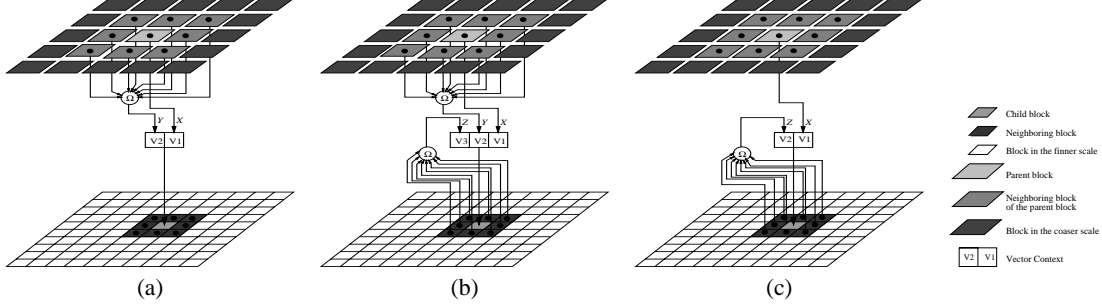
### 2.2. Three Context Structures

In [3], the context information was limited to the coarser scale, and resulted in poor boundary localization. This is because the boundary information may not be precisely exploited from the coarser scale. In practice, it was also found that only one context is not sufficient to achieve good segmentation results. In the following, we will develop three contexts to deal with these problems. We first introduce the ‘‘neighborhood’’ of  $D_{j,i,k}$  which is denoted by  $D_{j,i,k}^R$  to indicate the region of eight neighboring blocks of  $D_{j,i,k}$ . Then the class label of  $D_{j,i,k}^R$  is defined as  $C_{j,i,k}^R$  and obtained by

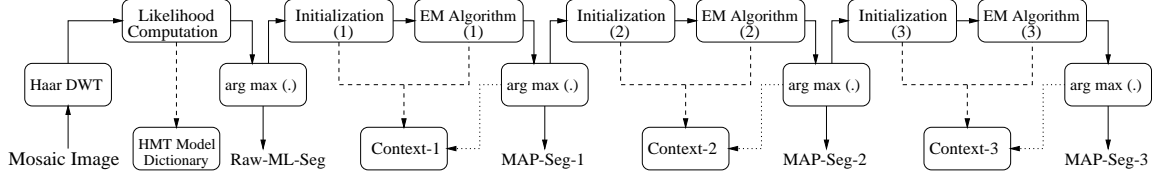
$$C_{j,i,k}^R = \arg \max_{c \in \{1, \dots, N_c\}} \left( \sum_{m=i-1}^{i+1} \sum_{n=k-1}^{k+1} p(C_{j,m,n} = c | \mathbf{D}, \mathbf{V}) - p(C_{j,i,k} = c | \mathbf{D}, \mathbf{V}) \right), \quad (4)$$

where the conditional pmf’s of eight neighboring blocks of  $D_{j,i,k}$  but not itself are used to determine  $C_{j,i,k}^R$ .  $C_{j,i,k}^R$  is able to provide some context information regarding to the class label of  $D_{j,i,k}$  or those of the child blocks of  $D_{j,i,k}$ .

The vector context  $V = \{V^{(1)}, \dots, V^{(d)}\}$  is determined by two factors, the dimension  $d$  and the value of each entry  $V^{(p)}$ ,  $p = 1, \dots, d$ , which takes value in  $\{1, \dots, N_c\}$ , thus we have  $N_v = (N_c)^d$ . Inspired by the context used in [3], we develop three new context structures shown in Fig. 2, and we assume that the context



**Fig. 2.** Three different context structures. (a) Context-1. (b) Context-2. (c) Context-3. The operation  $\Omega$  is defined in (4)



**Fig. 3.** The improved HMTseg algorithm using HCLT. The dashed line represents the association of a context structure in the EM algorithm and the initialization. The operation of  $\arg \max (\cdot)$  is defined in (2) where the conditioning is denoted by the dotted lines. The HMT model dictionary stores  $N_c$  many HMT model parameters. The “Raw-ML-Seg” denotes the initial ML segmentation results, and “MAP-Seg-1”, “MAP-Seg-2” and “MAP-Seg-3” represent the step-by-step results.

of  $D_{j,i,k}$  may be related to three variables defined as follows,

$$X = C_{j+1,x,y} \quad (5)$$

$$Y = C_{j+1,x,y}^R \quad (6)$$

$$Z = C_{j,i,k}^R \quad (7)$$

where  $x = \lfloor i/2 \rfloor$  and  $y = \lfloor k/2 \rfloor$ .  $X$  captures the effect from the parent block  $D_{j+1,x,y}$ ,  $Y$  characterizes the dependencies between the neighborhood of the parent block and  $D_{j,i,k}$ .  $Z$  exploits the relation between  $D_{j,i,k}$  and its neighborhood. From Fig. 2, we can determine the vector context  $V_{j,i,k}$ , by three different ways:

- Context-1:  $d = 2$ ,  $V_{j,i,k}^{(1)} = X$ ,  $V_{j,i,k}^{(2)} = Y$ ;
- Context-2:  $d = 3$ ,  $V_{j,i,k}^{(1)} = X$ ,  $V_{j,i,k}^{(2)} = Y$ ,  $V_{j,i,k}^{(3)} = Z$ ;
- Context-3:  $d = 2$ ,  $V_{j,i,k}^{(1)} = X$ ,  $V_{j,i,k}^{(2)} = Z$ .

Three context structures above have different advantages on the texture segmentation with respect to the classification of homogeneous regions and the localization of texture boundaries. We expect that Context-1 is more efficient for the classification of homogeneous regions, Context-2 has some balanced consideration between homogeneous regions and texture boundaries, and Context-3 is more suitable for the localization of texture boundaries. This statement has been verified by extensive experiments in practice.

### 2.3. Hybrid Contextual Labeling Tree

In practice, neither single context works well, and it is rational to integrate them together. Then, the problem is how to combine them in order to obtain good segmentation results in both homogeneous regions and texture boundaries simultaneously. Based on the analysis in Section 2.2, we develop a Hybrid Contextual Labeling Tree (HCLT) for the interscale fusion where the three context structures are cascaded serially in the order of Context-1,

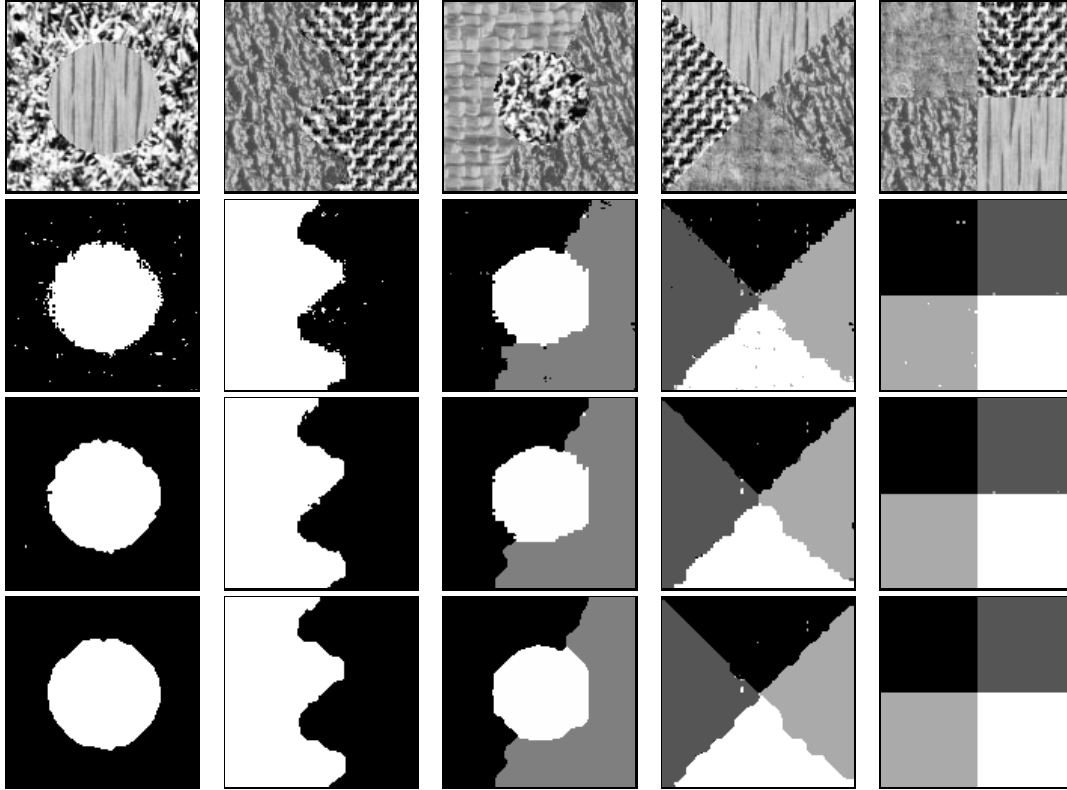
Context-2 and Context-3. Thus the EM training algorithm is performed based on the three contexts, respectively and sequentially, as shown in Fig. 3. It is worth noting that the vector context is recomputed for each dyadic block during each iteration of the EM algorithm to achieve the optimal training result.

We know that the initialization of the EM algorithm is often important to the efficiency and effectiveness of the training process. The initialization of the HCLT training is first associated with Context-1, and we assume that all blocks are initially independent. Given  $\mathbf{C}$  and  $\mathbf{D}$ , we use the countable statistics to estimate the initial setting of  $\alpha$ . Then the EM training algorithm is started based on Context-1 until it converges. Then we compute the initialization of the second EM training which is associated with Context-2, and so on. Directed by the three contexts, the EM training can progressively attain the desired solution.

## 3. SIMULATION RESULTS

The simulation of this work was conducted on a number of examples composed of  $128 \times 128$  texture mosaics taken from Brodatz album [4]. We use two criteria to evaluate the segmentation performance,  $PA_C$  which is the percentage of the accurate pixel classification, and  $PA_B$  which is the percentage of points on the localized the boundary that coincide with the true boundary.

We show the segmentation results for five mosaics in Fig. 4. We also tabulate the segmentation performances of  $PA_C$  and  $PA_B$  in Table 1. From Fig. 4 and Table 1, we can see that the new algorithm using the HCLT for the interscale fusions gives us good segmentation results in terms of both  $PA_C$  and  $PA_B$ . We also compare the improved HMTseg algorithm to some state-of-the-art algorithms in Table 2, where we can find that the new algorithm has the highest  $PA_C$  and the performance of  $PA_B$  is comparable to that in [5] where the boundary localization in the texture segmentation was investigated.



**Fig. 4.** The step-by-step segmentation results for 5 mosaics using the HCLT. The first row shows five texture mosaics, Mosaic1 (wood/grass), Mosaic2 (leather/weave), Mosaic3 (raffia/grass/leather), Mosaic4 (weave/wood/leather/cloth) and Mosaic5 (cloth/leather/weave/wood). The second row gives MAP-Seg-1. The third row presents MAP-Seg-2. The fourth row demonstrates the last MAP-Seg-3 results for 5 mosaics.

**Table 1.** Step-by-step results of HCLT.

Texture Mosaics	MAP-Seg-1		MAP-Seg-2		MAP-Seg-3	
	$PA_C$	$PA_B$	$PA_C$	$PA_B$	$PA_C$	$PA_B$
Mosaic1	97.5	33.8	98.7	64.0	99.2	77.8
Mosaic2	97.7	34.0	97.8	40.4	98.2	48.7
Mosaic3	97.4	48.2	97.7	53.7	98.4	67.1
Mosaic4	96.0	40.4	96.9	54.1	97.2	59.9
Mosaic5	99.8	81.5	100	97.3	100	100

**Table 2.** Performance comparison of texture segmentation.

References	$PA_C$	$PA_B$
Fig. 4 (HCLT)	97.2-100	48.7-100
Choi & Baraniuk [3]	93.9-99.7	24.8-65.8
Yhann & Young [5]	87.7-100	54.5-100
Kaplan [6]	92.7	-

#### 4. CONCLUSIONS

In this paper, we have improved the HMTseg algorithm proposed by Choi and Baraniuk by studying the role of the CLT for the Bayesian interscale fusion. Three context structures of different advantages have been developed and integrated into a hybrid CLT (HCLT) where the EM training is performed three times based on the three contexts respectively and sequentially, so that the seg-

mentation results can be refined step-by-step. The new algorithm improves the segmentation performances in both homogeneous regions and texture boundaries.

#### 5. REFERENCES

- [1] C. A. Bouman and M. Shapiro, "A multiscale random field model for bayesian image segmentation," *IEEE Trans. on Image Processing*, vol. 3, no. 2, pp. 162–177, March 1994.
- [2] M. S. Crouse, R. D. Nowak, and R. G. Baraniuk, "Wavelet-based statistical signal processing using hidden Markov models," *IEEE Trans. Signal Processing*, vol. 46, no. 4, pp. 886–902, April 1998.
- [3] H. Choi and R. Baraniuk, "Image segmentation using wavelet-domain classification," in *Proc. of SPIE*, Denver, CO, July 1999, vol. 3816, pp. 306–320.
- [4] P. Brodatz, *Textures—A Photographic Album for Artists and Designers*, New York: Dover, 1966.
- [5] S. R. Yhann and T. Y. Young, "Boundary localization in texture segmentation," *IEEE Trans. on Image Processing*, vol. 4, no. 6, pp. 849–855, June 1995.
- [6] L. M. Kaplan, "Extended fractal analysis for texture classification and segmentation," *IEEE Trans. on Image Processing*, vol. 8, no. 11, pp. 1572–1585, November 1999.