Analysis of Multiscale Texture Segmentation using Wavelet-Domain Hidden Markov Models

Hyeokho Choi, Brent Hendricks, and Richard Baraniuk *

Rice University, Department of Electrical and Computer Engineering, Houston, Texas 77005

Abstract

Wavelet-domain Hidden Markov Tree (HMT) models are powerful tools for modeling the statistical properties of wavelet transforms. By characterizing the joint statistics of the wavelet coefficients, HMTs efficiently capture the characteristics of a large class of real-world signals and images. In this paper, we apply this multiscale statistical description to the texture segmentation problem. We also show how the Kullback-Leibler (KL) distance between texture models can provide a simple performance indicator.

1 Introduction

The goal of an image segmentation algorithm is to assign a class label to each pixel of an image based on the properties of the pixels and their relationships with their neighbors. The segmentation process is a joint detection and estimation of the class labels and shapes of regions with homogeneous behavior.

For proper segmentation of images, both the large and small scale behaviors should be utilized to segment both large, homogeneous regions and detailed boundary regions. Thus, it is natural to approach the segmentation problem using multiscale analysis. Efforts have been exerted to model this multiscale behavior with autoregressive models [1, 2] and multiscale random fields [3]. In this paper, we propose a multiscale texture segmentation algorithm based on the wavelet transform.

Recently, the wavelet-domain Hidden Markov Tree (HMT) model was proposed to model the statistical properties of wavelet transforms [4, 5]. By modeling each wavelet coefficient as a Gaussian mixture density and by capturing the dependencies between wavelet coefficients as hidden state transitions, HMTs provide a natural setting for exploiting the structure inherent in real-world signals and images for signal detection and classification.

In this paper, we apply the tree structure of the HMT model to multiscale signal classification. By computing the likelihoods of dyadic sub-blocks of the image at different scales, we obtain several "raw" segmentations. Coarse scale segmentations are more reliable for large, homogeneous regions, while fine scale segmentations are more appropriate around boundaries between different textures. By combining raw segmentations from different scales, we obtain a robust and accurate overall result. In addition we demonstrate the use of the Kullback-Leibler (KL) distance between models as a preliminary performance indicator. Before we develop these new algorithms, we sketch some background on wavelets and wavelet-domain HMT models.

2 Background

2.1 The wavelet transform

The discrete wavelet transform (DWT) represents a 1-d signal $z(t)$ in terms of shifted versions of a lowpass scaling function $\phi(t)$ and shifted and dilated versions of a prototype bandpass wavelet function $\psi(t)$ [6]. For special choices of $\phi(t)$ and $\psi(t)$, the functions $\psi_{j,k}(t) \equiv 2^{j/2}\psi\left(2^j t - k\right)$, $\phi_{j,k}(t) \equiv 2^{j/2}\phi\left(2^j t - k\right)$, with $j, k \in \mathbb{Z}$ form an orthonormal basis, and we have the representation [6]

$$z \;=\; \sum_k u_{j_0,k}\,\phi_{j_0,k} \;+\; \sum_{j=j_0}^{\infty}\sum_k w_{j,k}\,\psi_{j,k}, \qquad (1)$$

with $u_{j,k} \equiv \int z(t)\,\phi^*_{j,k}(t)\,dt$ and $w_{j,k} \equiv \int z(t)\,\psi^*_{j,k}(t)\,dt$.

The *wavelet coefficient* $w_{j,k}$ measures the signal content around time $2^{-j}k$ and frequency $2^j f_0$. The scaling coefficient $u_{j,k}$ measures the local mean around time $2^{-j}k$. The DWT (1) employs scaling coefficients only at scale j_0; wavelet coefficients at scales $j > j_0$ represent higher resolution approximation to the signal. Any filter bank DWT implementation produces all of the scaling coefficients $u_{j,k}$, $j > j_0$ as a natural byproduct [6].

To keep the notation manageable in the sequel, we will adopt an abstract index scheme for the DWT coefficients: $u_{j,k} \to u_i, w_{j,k} \to w_i$.

*This work was supported by NSF, grant MIP–9457438, DARPA/AFOSR, grant F49620–97–1–0513, Texas Instruments, and the Rice Consortium for Computational Seismic Interpretation.

Email: choi@ece.rice.edu, richb@rice.edu, brentmh@rice.edu

Internet: www.dsp.rice.edu

We can easily construct 2-d wavelets from the 1-d ψ and ϕ by setting $\mathbf{x} \equiv (x, y) \in \mathbb{R}^2$ and $\psi^{\mathrm{HL}}(\mathbf{x}) = \psi(x)\phi(y)$, $\psi^{\mathrm{LH}}(\mathbf{x}) = \phi(x)\psi(y)$, and $\psi^{\mathrm{HH}}(\mathbf{x}) = \psi(x)\psi(y)$. If we let $\Psi \equiv \{\psi^{\mathrm{HL}}, \psi^{\mathrm{LH}}, \psi^{\mathrm{HH}}\}$, then the set of functions $\{\psi_{j,\mathbf{k}} \equiv 2^j \psi(2^j \mathbf{x} - \mathbf{k})\}_{\psi \in \Psi, j \in \mathbb{Z}, \mathbf{k} \in \mathbb{Z}^2}$ forms an orthonormal basis for $L_2(\mathbb{R}^2)$; that is, for every $z \in L_2(\mathbb{R}^2)$, we have

$$z = \sum_{j > j_0, \mathbf{k} \in \mathbb{Z}^2, \psi \in \Psi} w_{j,\mathbf{k},\psi}\, \psi_{j,\mathbf{k}} + \sum_{\mathbf{k} \in \mathbb{Z}^2} u_{j_0,\mathbf{k}}\, \phi_{j_0,\mathbf{k}}, \qquad (2)$$

with $u_{j_0,\mathbf{k}} \equiv \int_{\mathbb{R}^2} z(\mathbf{x}) \phi_{j_0,\mathbf{k}}(\mathbf{x}) d\mathbf{x}$ and $w_{j,\mathbf{k},\psi} \equiv \int_{\mathbb{R}^2} z(\mathbf{x}) \psi_{j,\mathbf{k}}(\mathbf{x}) d\mathbf{x}$.

2.2 Hidden Markov tree model

The *compression property* of the wavelet transform states that the transform of many real-world signals consists of a small number of large coefficients and a large number of small coefficients. We can consider the collection of small wavelet coefficients as outcomes of a probability density function (pdf) with small variance. Similarly, the collection of large coefficients can be considered as outcomes of a pdf with large variance. Hence, the pdf $f_{W_i}(w_i)$ of each wavelet coefficient is well approximated by *Gaussian mixture model*. To each wavelet coefficient W_i, we associate a discrete hidden state S_i that takes on values $m = 1, \ldots, M$ with probability mass function (pmf) $p_{S_i}(m)$. Conditioned on $S_i = m$, W_i is Gaussian with mean $\mu_{i,m}$ and variance $\sigma^2_{i,m}$. Thus, its overall pdf is given by

$$f_{W_i}(w_i) = \sum_{m=1}^{M} p_{S_i}(m)\, f_{W_i|S_i}(w_i|S_i = m). \qquad (3)$$

We consider only the case of $M = 2$ in this paper; however, the Gaussian mixture model can provide an arbitrarily close fit to the actual $f_W(w)$ as $M > 2$.

To generate a realization of W using the mixture model, we first randomly select a state variable S according to $p_S(s)$ and then draw an observation w according to $f_{W|S}(w|S = s)$. Although each wavelet coefficient W is conditionally Gaussian given its state variable S, the wavelet coefficient has an overall non-Gaussian density due to the randomness of S.

HMT models are multidimensional mixture models in which the hidden states have a Markov dependency structure [4]. Once we model the marginal density of each wavelet coefficient as a Gaussian mixture model, the correlation between wavelet coefficients can be captured by specifying the joint pmf of the hidden states. To capture the "persistence" of large/small values of wavelet coefficients across scales, we can model the correlations between wavelet coefficients as a binary tree where each branch indicates the dependency between the connected coefficients. Although coefficients that are not connected by the binary tree model are also correlated, we ignore these dependencies to simplify the model.

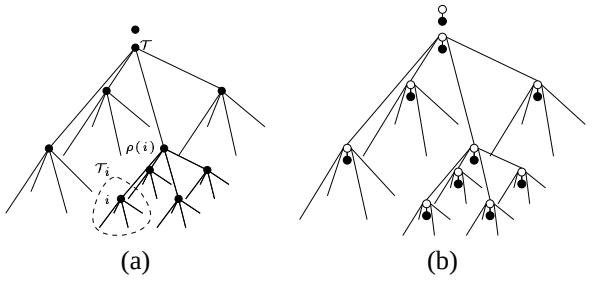


Figure 1. (a) Quadtree of 2-d wavelet coefficients for each subband, (b) 2-d wavelet-domain HMT model. We model each coefficient as a Gaussian mixture with a hidden state variable. Black nodes represent wavelet coefficients; white nodes represent hidden mixture state variables. Connecting the states vertically across scale yields the HMT model.

In order to describe the relationships between wavelet coefficients, we will use the notation $\rho(i)$ for the parent of node i. We also define $\mathcal{T}_i$ as the subtree of wavelet coefficients with root at node i, so that the subtree $\mathcal{T}_i$ contains coefficient w_i and all of its descendants.

The HMT model is specified via the Gaussian mixture parameters $\mu_{i,m}, \sigma^2_{i,m}$, the transition probabilities $\epsilon^{mn}_{i,\rho(i)} = p_{S_i|S_{\rho(i)}}(m|S_{\rho(i)} = n)$, and the pmf $p_{S_1}(m)$ for the root node S_1. These parameters can be grouped into a model parameter vector $\mathbf{\Theta}$. We train the HMT to capture the wavelet-domain characteristics of the signals of interest using the iterative Expectation Maximization (EM) algorithm [4]. For a given set of training signals, the trained model $\mathbf{\Theta}$ approximates the joint pdf $f(\mathbf{w})$ of all wavelet coefficients

In the HMT model, each wavelet coefficient W_i is conditionally independent of all other random variables given its state S_i. Furthermore, given the parent state $S_{\rho(i)}$, the nodes $\{S_i, W_i\}$ are independent of the entire tree except for S_i's descendants. The Markov structure of the model is on the states of the wavelet coefficients, not on the coefficients themselves (see Figure 1(b)).

The wavelet HMT model easily generalizes to 2-d using a quadtree model to capture the dependencies between the wavelet coefficients, with each wavelet state connected to the four "child" wavelet states below it (see Figure 1). The EM algorithm for the 1-d HMT model in [4] can be used without modification if we interpret the parent-child relations between nodes appropriately for quadtrees.

3 Multiscale Segmentation using HMT

3.1 Multiscale classification

The key step in our segmentation algorithm is to classify dyadic blocks of the image at different scales based on trained HMT models. For classification, we use the principle of maximum likelihood detection.[1]

Given an image of $2^J \times 2^J$ pixels, we define the *dyadic squares* at scale j to be the squares obtained by dividing the image into $2^j \times 2^j$ square regions of size $2^{J-j} \times 2^{J-j}$ pixels each for $j = 0, \ldots, J$. Denote each dyadic square as D_i, where i is an abstract index, with $J(i)$ the scale of D_i. In the sequel we will use the Haar wavelet transform. Because each Haar wavelet coefficient is computed from the pixel values in a dyadic square, we have a one-to-one correspondence between the wavelet coefficients and the dyadic squares.

Although the dyadic squares at a certain scale are correlated, we assume that they are independent for the multiscale classification step. We capture the dependencies later by combining classification results from different scales.

Texture classification using the HMT is simple: First, we obtain wavelet-domain HMT models for the candidate textures by training HMTs on hand-segmented training images. Then, to classify a dyadic block, we compute the conditional likelihood of the corresponding subtree for each candidate texture model; the texture model maximizing the likelihood is chosen as the texture of the block. This likelihood computation is easily implemented using the HMT EM algorithm [4].

In a 2-d HMT model $\mathbf{\Theta}$, we have three quadtrees corresponding to three different subbands. Denote the three quadtree models as $\mathbf{\Theta}^{\mathrm{HH}}$, $\mathbf{\Theta}^{\mathrm{HL}}$ and $\mathbf{\Theta}^{\mathrm{LH}}$, respectively. For subtree $\mathcal{T}_i^{\mathrm{HH}}$ in subband $\mathbf{\Theta}^{\mathrm{HH}}$ corresponding to dyadic square D_i, we compute the conditional likelihood $\beta_i(m) = f(\mathcal{T}_i^{\mathrm{HH}}|S_i = m, \mathbf{\Theta}^{\mathrm{HH}})$ and conditional probability $p(S_i = m|\mathbf{w}, \mathbf{\Theta}^{\mathrm{HH}})$, where $\mathbf{w} = \{w_i\}$ is the collection of all wavelet coefficients in the subband HH. Then, the conditional likelihood $f(\mathcal{T}_i^{\mathrm{HH}}|\mathbf{\Theta}^{\mathrm{HH}})$ can be computed as

$$f(\mathcal{T}_i^{\mathrm{HH}}|\mathbf{\Theta}^{\mathrm{HH}}) = \sum_{m=1}^{M} \beta_i(m) p(S_i = m|\mathbf{w}, \mathbf{\Theta}^{\mathrm{HH}}). \quad (4)$$

For the HL and LH subbands, we can similarly compute the likelihoods $f(\mathcal{T}_i^{\mathrm{HL}}|\mathbf{\Theta}^{\mathrm{HL}})$ and $f(\mathcal{T}_i^{\mathrm{LH}}|\mathbf{\Theta}^{\mathrm{LH}})$, respectively.

There are several ways in which we can combine the likelihoods of the wavelet coefficients from different subbands. The simplest and most effective method we have found assumes that all three subbands are independent. Then, we simply multiply the three likelihood functions together to obtain the total likelihood $f(D_i|\mathbf{\Theta})$ of the dyadic square D_i. Another possibility ties the three root nodes of the quadtrees together and compute likelihoods for the combined tree.

For each node i, the dyadic square D_i can be classified by finding the texture model $\mathbf{\Theta}$ for which the likelihood $f(D_i|\mathbf{\Theta})$ is maximized, producing a raw segmentation of the image down to 2×2 pixels image blocks. We call this segmentation a "raw" segmentation, because we do not use any relationship between segmentations across different scales. We expect this "raw" segmentation to be more reliable at coarse scales, because classification of large image blocks is more reliable due to their richness in statistical information. However, at coarse scales, the boundaries between different textures will not be captured accurately. At fine scales, the segmentation is less reliable, but boundaries are better captured.

When we compute the likelihoods of wavelet coefficient subtrees, we ignore the scaling coefficients. Thus, we do not take advantage of the information in the local brightness of the image. This is why we can classify only down to 2×2 blocks. We can obtain a pixel-level classification using the histogram of pixel brightness for each texture.

However, because we ignore the scaling coefficients, the local brightness levels of the texture regions do not affect the performance down to 2×2 scale. This is an advantage of the use of HMT model for texture segmentation. The HMT model captures only joint statistics between pixels, and thus it is robust to the change of local average brightness. This is a desirable feature because the local brightness of an image often varies at different regions. We now fuse the results from each scale to obtain a final, reliable segmentation.

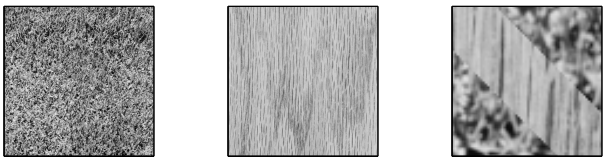
4 Examples

In the first set of simulations, we segment a simple synthesized texture image consisting of a combination of "grass" and "wood" texture images from the USC Image Database.[2] To keep things simple, we considered only 64×64 images in the simulation. First we randomly selected five 64×64 blocks from the original "grass" and "wood" images to train respective HMT models. The training was performed with intra-scale tying [4] to avoid overfitting. The test image was created from randomly chosen 64×64 grass and wood images. The mosaic test image is shown in Fig. 2(c); the (i, j) elements with $|i - j| \leq 18$ correspond to wood texture and the remaining region is grass texture.

Figure 3 shows the segmentation results at multiple scales. At coarse scales, the classification is reliable, but the details at the texture boundaries are not well represented due

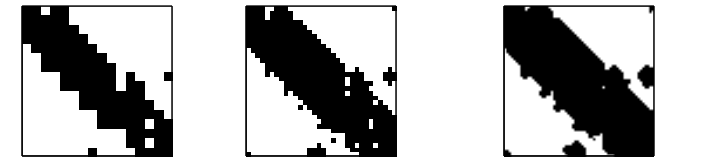
[1] We can also use the maximum a posteriori probability detection if we have prior pmfs of texture classes.

[2] `http://sipi.usc.edu/services.html`



(a) (b) (c)

Figure 2. (a) Grass and (b) wood texture images from the USC image data base. (c) 64×64 mosaic test image to be segmented.



4×4-block scale 2×2-block scale pixel-level segmentation

Figure 3. Segmentation results for grass-wood image

to the large block size. The boundaries are better classified at finer scales, but we make more classification errors because of the paucity of statistical information in each small block.

For the second example, we trained HMT and pixel brightness models for "text," "image," and "background" textures using hand-segmented blocks from the 512×512 document in Fig. 4(a). Choosing $j_0 = 3$ for the starting scale (corresponding to 6-scale quad-trees on 64×64 image blocks), we segmented the 512×512 test image in Fig. 4(b).

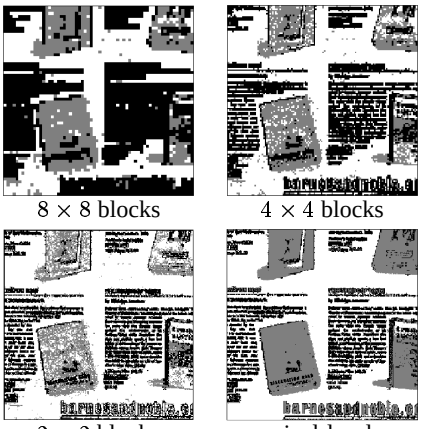
Fig. 4(c) illustrates the resulting segmentation. Text, image, and background regions are displayed as black, gray, and white, respectively. All text regions were segmented well, including the text surrounded by images on the books. At the bottom, we observe that the large-font title text was segmented as image. This is because the homogeneous texture inside each large letter had properties more similar to images than (small-font) text. The background regions were correctly segmented, even though the brightness of the background varies in different areas and is corrupted by a noise-like feature caused by text on the reverse side of the page.

5 Accuracy Assessment using KL distance

Our image segmentation technique is highly dependent upon accurate classification of the dyadic squares. A question which arises naturally is how to determine the quality of this classification step. We examine the use of the Kullback-Leibler distance as a quality indicator.

The KL distance between pdf's f^1 and f^0 is expressed

(a) training image (b) test image



8×8 blocks 4×4 blocks

2×2 blocks pixel-level

(c) interscale fused segmentations

Figure 4. Document segmentation using HMTseg.

as:

$$D(f^0||f^1) = \int_{\Re^n} f^0(\vec{x}) \ln \frac{f^0(\vec{x})}{f^1(\vec{x})}\, d\vec{x} \qquad (5)$$

While not a true distance metric, it does give a us sense of how "far" one density is from another. This intuitive sense of distance is made more rigorous by examining Stein's lemma, which states that for binary hypothesis testing between p^1 and p^0,

$$\text{If } P_M \leq \alpha, \text{ then } \lim_{L\to\infty} -\frac{\ln P_F}{L} = D(p^1||p^0) \qquad (6)$$

where P_M is the probability of guessing hypothesis 0 when hypothesis 1 is true, and vice versa for P_F. This tells us that the error probability decays at best exponentially with the KL distance between the two hypotheses.

Rather than use a numerical integration technique, we approximate the KL distance by treating it as an expected value, since

$$D(f^0||f^1) = E_{f^0}\left[\ln \frac{f^0(\mathbf{x})}{f^1(\mathbf{x})}\right] \qquad (7)$$

We now apply known statistical techniques for mean estimation. Estimating the KL distance in this way consists of three separate steps.

1. *Generating data.* First, we generate a large number of realizations (data sequences) under model 0. This step is accomplished in a straightforward manner by first assigning states down the the tree according to the state transitions. Given these states, we then randomly draw the coefficients themselves from the correct distribution.

2. *Computing likelihoods.* The likelihood of the data is easily obtained by using the EM algorithm [4]. In our case, we take the data generated by the first model, and compute the difference of log likelihoods of the data under each of the two models. This gives us the $\ln \frac{f^0(\mathbf{x})}{f^1(\mathbf{x})}$ term from (7), which we calculate for each of the data sequences in step 1.

3. *Computing statistics.* The estimate for the KL distance is obtained by averaging the log likelihood values found in the previous step. Here we invoke the law of large numbers (LLN) to claim that as the number of samples increases (note: this refers to the *number* of data sequences, not the *size* of the data sequences), the empirical average will approach the expected value.

The KL distance results for the models in the above examples are listed in Table 1. It is important to note that the KL distance only yields an indication of classification accuracy if the models are well matched to the data. It has been shown [7] that this is true for a large class of real-world images.

If the models do not closely match the data, then the KL distance may not give an indication of the the true misclassification rate. Without any a priori knowledge of the fitness of the models, a large KL distance yields no information about segmentation performance. On the other hand, a small KL distance does indicate that either the model is a poor fit, or that the models match the data well, but are too close probabilistically to perform a good classification. In either case, a small KL distance predicts a poor performance by the HMT segmentation algorithm.

6 Conclusions

In this paper, we have developed a new framework for texture segmentation based on wavelet-domain hidden Markov models. By concisely modeling the statistical behavior of textures at multiple scales and by combining segmentations at multiple scales, the algorithm produces a robust and accurate segmentation of texture images.

We believe the proposed segmentation algorithm can be applied to many different types of images, including radar/sonar images, medical images (CT/ultrasound), and document images (text/picture segmentation). To incorporate different characteristics of various images, the use of different wavelet bases and more complicated context models should be investigated. The KL distance is a good initial predictor of segmentation performance, but generally the analysis of detection errors for wavelet-domain HMT based classifiers remains a future research topic.

Table 1: KL distances between HMT models

$D(f_{grass} \| f_{wood})$			
Samples	10	100	1000
Mean	**17711**	**17610**	**17712**
Bias	-9.725	-4.718	-0.581
Std. error	238.98	108.03	28.64
$D(f_{figure} \| f_{background})$			
Samples	10	100	1000
Mean	**8570**	**9014**	**9158**
Bias	107.1	-47.25	-2.54
Std. error	916.7	348.3	108.6
$D(f_{text} \| f_{background})$			
Samples	10	100	1000
Mean	**40696**	**41195**	**41021**
Bias	2.75	16.40	5.044
Std. error	447.0	139.66	47.28
$D(f_{figure} \| f_{text})$			
Samples	10	100	1000
Mean	**2056**	**2081**	**2093**
Bias	1.78	0.456	-0.156
Std. error	29.97	9.386	3.204

References

[1] M. Basseville et al., "Modeling and Estimation of Multiresolution Stochastic Processes," *IEEE Trans. Inf. Theory*, vol. 38, no. 2, pp. 766-784, Mar. 1992.

[2] C. H. Fosgate et al., "Multiscale Segmentation and Anomaly Enhancement of SAR Imagery," *IEEE Trans. Image Proc.*, vol. 6, no. 1, pp. 7-20, Jan. 1997.

[3] C. A. Bouman and M. Shapiro, "A Multiscale Random Field Model for Bayesian Image Segmentation," *IEEE Trans. Image Proc.*, vol. 3, no. 2, pp. 162-177, Mar. 1994

[4] M. S. Crouse, R. D. Nowak, and R. G. Baraniuk, "Wavelet-Based Statistical Signal Processing Using Hidden Markov Models," *IEEE Trans. Signal. Proc.*, vol. 46, no. 4, pp. 886-902, Apr. 1998.

[5] M. S. Crouse and R. G. Baraniuk, "Contextual Hidden Markov Models for Wavelet-domain Signal Processing," in *Proc. 31st Asilomar Conference*, vol. 1, pp. 95-100, Nov. 1997.

[6] I. Daubechies, *Ten Lectures on Wavelets.* New York: SIAM, 1992.

[7] J. Romberg, H. Choi, and R. Baraniuk, "Bayesian Tree-Structured Image Modeling using Wavelet-domain Hudden Markov Models," *Proc. SPIE*, pp. 31-44, July 1999.