

A New Wavelet-Domain HMTseg Algorithm for Remotely Sensed Image Segmentation

Qiang Sun, Biao Hou, and Li-cheng Jiao

Institute of Intelligent Information Processing, Xidian University,
710071 Xi'an, China
art_navigator@yahoo.com.cn

Abstract. A new wavelet-domain HMTseg method is proposed, which fuses the segmentation results at coarse and fine scales with a new and feasible context model together with one preprocessing of raw segmentations at different scales. Compared to the original HMTseg method, the new method not only lays emphasis on the performance from coarse-scale segmentation, preserves the main outlines of the homogeneous regions in an image, and thus achieves good region consistency of segmentation, but also take into account the information from fine-scale segmentation, thus improves the accuracy of boundary localization of segmentation and enables the discrimination of small targets in an image, which is desirable for interpretation of remotely sensed images. Experiments on remotely sensed images, including aerial photos and SAR images, demonstrate that the proposed method can effectively take into consideration both the region consistency and the accuracy of boundary localization of segmentation performance, and give better segmentation results.

1 Introduction

In most recent years, wavelet-domain statistical image models, especially hidden Markov tree (HMT) [1] models, have gained more and more attention from image processing and analysis community due to their effectiveness and flexibility in performing image analysis tasks. Choi et al proposed a new framework, HMTseg [2], for multiscale Bayesian image segmentation based on wavelet-domain HMT models pioneered by Crouse et al to give the statistical characterization of signals by capturing inter-scale dependencies of wavelet coefficients. In HMTseg method, the raw maximum likelihood (ML) segmentations at different scales are yielded before the inter-scale fusion of class labels from coarse scale to fine one. Another tree structure, the context labeling tree (CLT), was designed to exploit the dependencies of parent and child labels for the dyadic image squares across scales. The final classification of each dyadic square at different scales, except for the coarsest one of wavelet decomposition, was implemented in manner of scale recursion. In [2], one simplified context model is used to implement the inter-scale fusion in HMTseg method. This model is typically effective for images mostly made up of homogeneous regions, but fails to perform well for images consisting of complex structured information and/or

more inhomogeneous regions, remotely sensed images involved in this paper for example. In addition, this segmentation method, suitable for natural or textured images, could not give good results for synthetic aperture radar (SAR) images due to a particular kind of noise, speckle, inherent in them.

In this paper, we propose a modified HMTseg method to consider the region consistency (robust classification) of ML raw segmentation for dyadic squares at coarse scale and the accuracy of boundary localization (poor classification) at fine scale, based on a new and feasible context model taking into account the information from coarser scales and fine ones simultaneously. Meanwhile, a preprocessing stage is also introduced to further amend the raw ML segmentations at different scales to favor better multiscale fusion eventually.

This paper is organized as follows. The HMT method [2] is briefly reviewed in section 2. In section 3, the image segmentation using the modified HMTseg method is detailed. Simulation results and analysis are given in section 4. Finally, one conclusion is drawn and future work directed in section 5.

2 HMTseg Method for Multiscale Image Segmentation

The HMTseg method relies on three separate tree structures: the wavelet transform quad-tree, the HMT, and a labeling tree [2]. As a complete procedure, it includes three essential ingredients, i.e. HMT model training, multiscale likelihood computation, and fusion of multiscale maximum likelihood (ML) raw segmentations.

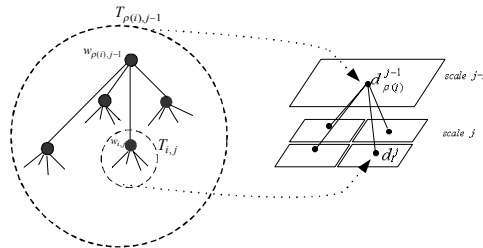


Fig. 1. Correspondence of quad-tree structure of wavelet coefficients with multiscale representation of an image

HMT model training is used to fitting a certain set of model parameters to a given set of training data extracted from a certain homogeneous region in an image. To this end, an iterative procedure, 2-D extension of EM algorithm for 1-D HMT model [1], is exploited to find one locally optimal set Θ_c of model parameters for a given set of training data. The likelihood that training data come from a certain model is maximized via a few iterations of E and M steps of the EM algorithm, as a consequence the set Θ_c of model parameters that locally maximizes the likelihood can be obtained. To train a pixel brightness probability density function (PDF), Gaussian mixture model (GMM) is used to train each texture and thus to obtain pixel-level segmentations.

As for multiscale likelihood computation, a single upward sweep [a fast $O(n)$ algorithm] through the HMT can deal with it [2]. Each subtree of wavelet coefficients residing in one of the subbands corresponds to one specific dyadic image square at each different scale. For example, the subtree $T_{i,j}$ rooted at $w_{i,j}$ can be brought into correspondence with the dyadic square d_i^j , as illustrated in Fig. 1. Thus, the calculation of likelihood that the dyadic square d_i^j is produced by the parametric set Θ_c can be then transformed into the computation of the likelihood that the coefficient subtree $T_{i,j}$ is generated by Θ_c . In this way, the likelihoods of all dyadic squares of an image can be obtained in a single upward sweep through the tree of wavelet coefficients since the wavelet transform and HMT possess the similar multiscale structure, as detailed in [2]. In fact, the wavelet coefficients corresponding to the dyadic square d_i are composed of one triple $\{\mathcal{T}_i^{LH}, \mathcal{T}_i^{HL}, \mathcal{T}_i^{HH}\}$, each a subtree of one of the three wavelet subband quadrees. Under the independent subband assumption, the likelihood that the dyadic square d_i is gained by the model Θ_c can be expressed as

$$f(d_i|\Theta_c) = f(\mathcal{T}_i^{LH}|\Theta_c^{LH})f(\mathcal{T}_i^{HL}|\Theta_c^{HL})f(\mathcal{T}_i^{HH}|\Theta_c^{HH}). \tag{1}$$

Finally, the ML classification can be obtained by

$$\hat{c}_i^{ML} := \arg \max_{c \in \{1,2,\dots,N_c\}} f(d_i|\Theta_c), \tag{2}$$

where N_c is the number of texture classes in an image. For different scale representation of an image in a pyramidal structure, a set of ML segmentations can be obtained in the same way. Hence, this step yields the ML raw segmentations at different scales, which is the basis of the following fusion procedure.

The third step in HMTseg method is to intelligently combine the raw multiscale ML segmentations given at the second step using a Bayesian inter-scale fusion technique. This idea is based on the fact that finer-scale dyadic squares nest inside coarser-scale squares, and the dyadic squares are statistically dependent across scale for images consisting of fairly large, homogeneous regions [2]. During the fusion procedure, relatively reliable coarser-scale information is used to guide less reliable fine-scale decisions. This Bayesian inter-scale decision fusion computes one maximum a posteriori (MAP) estimate \hat{c}_i^{MAP} for the class label of each d_i , i.e.

$$\hat{c}_i^{MAP} := \arg \max_{c_i \in \{1,2,\dots,N_c\}} P(c_i|\mathbf{x}) = \arg \max_{c_i \in \{1,2,\dots,N_c\}} \frac{f(\mathbf{x}|c_i)P(c_i)}{f(\mathbf{x})}. \tag{3}$$

Let $\mathbf{d}^j := \{d_i^j\}$ denote the collection of all dyadic squares at scale j (each d_i^j contains complete information on the image \mathbf{x}). A posterior equivalent to (3) is then

$$\hat{c}_{i,j}^{MAP} := \arg \max_{c_i^j \in \{1,2,\dots,N_c\}} \frac{f(\mathbf{d}^j|c_i^j)P(c_i^j)}{f(\mathbf{d}^j)}, \tag{4}$$

where $\frac{f(\mathbf{d}^j | \mathbf{c}_i^j) p(\mathbf{c}_i^j)}{f(\mathbf{d}^j)} = p(\mathbf{c}_i^j | \mathbf{d}^j)$ is the marginal of the joint PDF $p(\mathbf{c}^j | \mathbf{d}^j)$ denoted as

$$p(\mathbf{c}^j | \mathbf{d}^j) = \frac{f(\mathbf{d}^j | \mathbf{c}^j) p(\mathbf{c}^j)}{f(\mathbf{d}^j)} = \frac{p(\mathbf{c}^j)}{f(\mathbf{d}^j)} \prod_i f(\mathbf{d}_i^j | \mathbf{c}_i^j). \quad (5)$$

Equation (5) is based on the assumption that each \mathbf{d}_i is generated with the distribution $f(\mathbf{d}_i | \mathbf{c}_i)$ independently of all other class labels \mathbf{c}_k and dyadic squares \mathbf{d}_k , $k \neq i$.

However, the calculation of the marginal of $p(\mathbf{c}^j | \mathbf{d}^j)$ above is generally a difficult task unless $p(\mathbf{c}^j)$ has a special structure. In [2], a new organization structure, context labeling tree (CLT), was used to alleviate this difficulty. In this structure, the joint distribution of all the class labels C_i^j (the random variable corresponding to \mathbf{c}_i^j) at scale j is completely determined by C_i^{j-1} at the immediately coarser scale and thus a Markov chain $\{C_i^{j-1}\} \rightarrow C_i^j \rightarrow \mathbf{d}_i^j$ is formed. Given $\mathbf{C}^{j-1} = \mathbf{c}^{j-1}$, the C_i^j s at scale j are independent, and the multiscale priori $p(\mathbf{c}^j | \mathbf{c}^{j-1}) = \prod_i p(\mathbf{c}_i^j | \mathbf{c}^{j-1})$ holds. However, due to the high dimensionality of the conditioning vector \mathbf{c}^{j-1} , the estimate of the marginalized class priori distribution $p(\mathbf{c}_i^j | \mathbf{c}^{j-1})$ still requires a prohibitive amount of training data [2]. In practice, a context vector \mathbf{v}_i^j , the function of the \mathbf{c}^{j-1} , is introduced to provide further simplification of the priori $p(\mathbf{c}^j | \mathbf{c}^{j-1})$. A notation \mathbf{v}^j is used to denote the collection of all contexts at scale j . Conditioned on the context \mathbf{v}^j , equation (5) can be reformulated as

$$p(\mathbf{c}^j | \mathbf{d}^j, \mathbf{v}^j) = \frac{f(\mathbf{d}^j | \mathbf{c}^j) p(\mathbf{c}^j | \mathbf{v}^j)}{f(\mathbf{d}^j | \mathbf{v}^j)} = \frac{1}{f(\mathbf{d}^j | \mathbf{v}^j)} \prod_i [f(\mathbf{d}_i^j | \mathbf{c}_i^j) p(\mathbf{c}_i^j | \mathbf{v}_i^j)]. \quad (6)$$

Thereby, the marginal $p(\mathbf{c}_i^j | \mathbf{d}_i^j, \mathbf{v}_i^j) \propto f(\mathbf{d}_i^j | \mathbf{c}_i^j) p(\mathbf{c}_i^j | \mathbf{v}_i^j)$ is obtained, a simplified version of the MAP $p(\mathbf{c}_i | \mathbf{x})$ in Eq. (3). The factor $f(\mathbf{d}_i^j | \mathbf{c}_i^j)$ has been computed in the second step of HMTseg method, and the conditional probability $p(\mathbf{c}_i^j | \mathbf{v}_i^j)$ can be denoted as

$$p(\mathbf{c}_i^j | \mathbf{v}_i^j) = \frac{f(\mathbf{v}_i^j | \mathbf{c}_i^j) p(\mathbf{c}_i^j)}{\sum_{\mathbf{c}_i^j=1}^{N_c} f(\mathbf{v}_i^j | \mathbf{c}_i^j) p(\mathbf{c}_i^j)}. \quad (7)$$

A new EM algorithm for CLT has been developed in [2] to solve for $f(\mathbf{v}_i^j | \mathbf{c}_i^j)$ and $p(\mathbf{c}_i^j)$. Thus far, the MAP estimate $\hat{\mathbf{c}}_i^{\text{MAP}}$ for the class label of each dyadic square \mathbf{d}_i has been on hand.

3 Modified Image Segmentation Using Proposed Method

Effective modeling of context models for each dyadic square d_i is crucial to effectively fuse the ML raw segmentations from coarse scale to fine one in order to obtain good results in multiscale fusion stage. In the original HMTseg method [2], the context v_i^j is specified as a vector of two entries consisting of the value of class label $C_{\rho(i)}$ of the parent square and the majority vote of the class labels of the parent plus its eight neighbors, as illustrated in Fig. 2 (a). This simplified context is typically effective for images made up of separate large homogeneous textures since it lays more emphasis on the information of class labels at coarse scales. However, the segmentation results are mostly unsatisfactory when the images in hand, remotely sensed images for example, include more complex structures. Therefore, one slightly complicated and yet feasible context model, as shown in Fig. 2(b), is introduced here to incorporate both the information about the class labels at the coarse scale and that at the fine scale so as to take into account the region consistency and edge accuracy of segmentation performance simultaneously, which will be detailed in section 3.2.

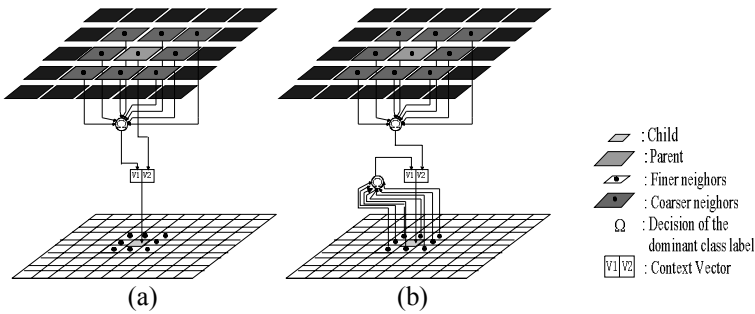


Fig. 2. (a) Context in [2] (b) new context proposed

3.1 Preprocessing of Raw Segmentations

As mentioned in section 2, the second step in the HMTseg method is multiscale likelihood computation by which the raw ML segmentations can be obtained. These raw segmentations are given only at the each individual scale without taking into the interactions across scales. Thus, the results from each single scale are typically unsatisfactory, which can be solved by the fusion of raw ML segmentations across scales using the third step in section 2. On the one hand, one appropriate context model is prerequisite to effectively implement fusion. On the other hand, modest preprocessing of raw segmentations could facilitate the following fusion and further improve segmentation performance. To this end, an 8-connectedness labeling [4] is introduced to amend the raw segmentations, which would favor better fusion results. This stage is carried out at each scale successively except for the coarsest scale of wavelet decomposition.

3.2 Multiscale Fusion Based on a New Context Model

Herein, a new context \mathbf{v}_i^j for dyadic square \mathbf{d}_i^j is defined, which consists of the majority vote of the class labels for the parent’s eight neighbors and that for the child’s eight neighbors.

The purpose of multiscale fusion is to maximize the conditional probability $p(c_i^j | \mathbf{d}_i^j, \mathbf{v}_i^j)$ with which the label \hat{c}_i^j for \mathbf{d}_i^j can be found with MAP criterion, i.e.

$$\hat{c}_i^j = \arg \max_{c \in \{1, 2, \dots, N_c\}} p(c_i^j | \mathbf{d}_i^j, \mathbf{v}_i^j) = \arg \max_{c \in \{1, 2, \dots, N_c\}} f(\mathbf{d}_i^j | c_i^j) p(c_i^j | \mathbf{v}_i^j), \tag{8}$$

where the computation of likelihood function $f(\mathbf{d}_i^j | c_i^j)$ has been completed in raw segmentation stage. Hence, the key task is to calculate the conditional probability $p(c_i^j | \mathbf{v}_i^j)$ based on which \hat{c}_i^j can be obtained. According to Eq. (7), the probabilities $f(\mathbf{v}_i^j | c_i^j)$ and $p(c_i^j)$ are undetermined temporarily. Here, we use the EM algorithm [2] specific for CLT to solve for $f(\mathbf{v}_i^j | c_i^j)$ and $p(c_i^j)$. Similar to the settings in [2], $f(\mathbf{v}_i^j | c_i^j)$ and $p(c_i^j)$ are assumed the same at each individual scale, and two notations $e_{j,m} := p(c_i^j = m)$, $\alpha_{j, \bar{\mathbf{v}}_m, n} := p(\mathbf{v}_i^j = \bar{\mathbf{v}}_m | c_i^j = n)$, $n \in \{1, \dots, N_c\}$, $m \in \{1, \dots, N_c\}$ are defined for all i at scale j . The EM algorithm to calculate $\mathbf{P} = \{e_{j,m}, \alpha_{j, \bar{\mathbf{v}}_m, n}\}$ runs as follows.

Initialize: Set counter $I=0$ and choose appropriate \mathbf{P}^0 ;

Expectation (E) step: Given \mathbf{P}^I , calculate

$$p(c_i^j = n | \mathbf{v}_i^j, \mathbf{d}_i^j) = \frac{e_{j,n} \alpha_{j, \mathbf{v}_i^j, n} f(\mathbf{d}_i^j | c_i^j = n)}{\sum_{c=1, \dots, N_c} e_{j,c} \alpha_{j, \mathbf{v}_i^j, c} f(\mathbf{d}_i^j | c_i^j = c)} ;$$

Maximization (M) step: Update the two elements of \mathbf{P}^{I+1}

$$e_{j,n} = \frac{1}{2^{2j}} \sum_i p(c_i^j = n | \mathbf{v}_i^j, \mathbf{d}_i^j) ,$$

$$\alpha_{j, \bar{\mathbf{v}}_m, n} = \frac{1}{2^{2j} e_{j,n}} \sum_{i \text{ with } \mathbf{v}_i^j = \bar{\mathbf{v}}_m} p(c_i^j = n | \mathbf{v}_i^j, \mathbf{d}_i^j) ;$$

Iterate: Increment $I \rightarrow I+1$, and apply E and M steps until converged.

3.3 Pixel-Level Segmentation

Pixel-level segmentation can not be obtained directly from the ML raw segmentation procedure, since the wavelet transform (Haar wavelet base is adopted in this paper) characterizes the joint statistics of dyadic squares only down to 2×2 blocks [2]. As mentioned in section 2, the GMM can be exploited to model the pixel intensity values for each training texture, based on which the likelihood for each pixel is calculated and the multiscale fusion algorithm above can be naturally extended to the pixel level, and the final segmentation is accomplished. The segmentation of SAR images, however, can not be performed as well as natural textured images due to the intrinsic speckle in them. We use here the truncated HMTseg method [3] in which a scale threshold J was chose so that only the coefficients corresponding to dyadic squares not more than d^J were trained. Moreover, a combination strategy, using HMT-based raw segmentations and pixel-intensity-based ones, was utilized during inter-scale fusion procedure to finally get better results.

4 Experimental Results and Analysis

The experiments were conducted on remotely sensed images including an aerial photo (256×256 pixels, 256 gray levels) from USC-SIPC image database [5] and an SAR image (256×256 pixels, 256 gray levels, China Lake Airport, California, 3-m resolution) from Sandia national laboratories SAR image repository [6], shown in Fig. 3 (a) and (c). The HMT models for different types of textures (two classes for the aerial photo and three classes for the SAR image) were firstly trained based on the training data with size of 64×64 manually extracted from two original images. The number of wavelet decomposition levels was restricted to four with the balance between time consumption for model training and the reliability of segmentations in mind. The pixel-level segmentations were performed using GMM technique.

The final fusion results using original HMTseg method [2] and our method are demonstrated in Fig. 4. Not only is region consistency of segmentation performance obtained but the accuracy of boundary localization is improved further in Fig. 4 (b) and (d) obtained using the proposed method. For example, the ports in the aerial photo and the runways encircling the airports in the SAR image can be mostly discriminated. The results suggest that it is important to make full use of the information from coarse and fine scales simultaneously in the fusion process to gain better segmentation results.

5 Conclusion and Future Work

In this paper, a modified HMTseg method is proposed using a new context model and a preprocessing stage introduced to further amend the raw ML segmentations at different scales to facilitate the final multiscale fusion. Based on the proposed method, performance for remotely sensed images are improved, especially in the accuracy of boundary localization. Edges in an image provide important information

for identifying the objects in remotely sensed images. Combining the edge cues in an image to devise an edge-guided segmentation method is our further work.

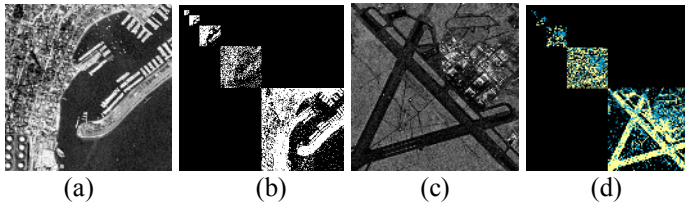


Fig. 3. Multiscale raw segmentation and pixel-level segmentation results of remotely sensed images: (a) aerial photo; (b) 4-level raw segmentation and pixel-level segmentation of (a); (c) SAR image; (d) 4-level raw segmentation and pixel-level segmentation of (c)

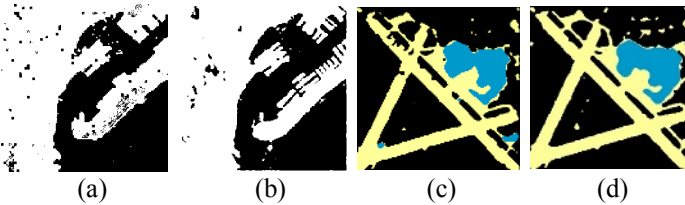


Fig. 4. Multiscale fusion results: (a) fusion result of Fig. 3 (b) by the method in [2]; (b) fusion result of Fig. 3 (b) by the proposed method; (c) fusion result of Fig. 3 (d) by the method in [2]; (d) fusion result of Fig. 3 (d) by the proposed method

References

1. Crouse, M.S. , Nowak, R.D. , Baraniuk, R.G.: Wavelet-Based Statistical Signal Processing Using Hidden Markov Models. *IEEE Trans. on Signal Processing.* 46 (1998) 886–902
2. Choi, H. , Baraniuk, R.G.: Multiscale Image Segmentation Using Wavelet-Domain Hidden Markov Models. *IEEE Trans. on Image Processing.* 10 (2001) 1309–1321
3. Venkatachalam, V. , Choi, H. , Baraniuk, R.G.: Multiscale SAR Image Segmentation Using Wavelet-Domain Hidden Markov Tree Models. In *Proc. of SPIE*, 4053 (2000) 1605–1611
4. Jain, A.K.: *Fundamentals of Digital Image Processing.* Prentice-Hall, 1989
5. The USC-SIPc Image Database. Available: <http://sipi.usc.edu/services.html>
6. Sandia Synthetic Aperture Radar Imagery Repository. Available: <http://www.sandia.gov/radar/imagery.html>