A STUDY OF CONTEXTUAL MODELING AND TEXTURE CHARACTERIZATION FOR MULTISCALE BAYESIAN SEGMENTATION

Guoliang Fan and Xiaomu Song
School of Electrical and Computer Engineering
Oklahoma State University, Stillwater, OK 74078
Email: {gfan,xiaos}@okstate.edu

ABSTRACT

In this paper, we demonstrate that multiscale Bayesian image segmentation can be enhanced by improving both contextual modeling and statistical texture characterization. Firstly, we show a joint multi-context and multiscale approach to achieve more robust contextual modeling by using multiple context models. Secondly, we study statistical texture characterization using wavelet-domain Hidden Markov Models (HMMs), and in particular, we use an improved HMM, HMT-3S, to obtain more accurate multiscale texture characterization. Experimental results show that both of them play important roles in multiscale Bayesian segmentation.

1. INTRODUCTION

Multiscale Bayesian approaches to image segmentation have been proven efficient to integrate both image features and prior contextual properties. In particular, in [1], Markovian dependencies are assumed across scales to capture interscale dependencies of multiscale class labels with a causal Markov Random Field (MRF) structure, so that a non-iterative segmentation algorithm was developed where a sequential MAP (SMAP) estimator was developed with the low computational cost. In [2], a trainable multiscale context model was introduced to characterize complex aspects of both local and global contextual behavior by off-line context model training. In [3], an efficient multiscale segmentation method was proposed to implement the interscale fusion of multiscale classification results by on-line context model training. These methods adopt SMAP estimators that mainly consider a single context model by capturing interscale dependencies of multiscale class labels. In [4], it was discussed that contextual modeling of multiscale class labels is critical to segmentation results, and a joint multi-context and multiscale (JMCMS) approach was proposed. The JMCMS uses multiple context models of distinct advantages sequentially and individually for Bayesian inference, and it outperforms the methods using a single context model, e.g., [3], at the comparable computational cost. In addition, statistical texture characterization is another important issue in Bayesian segmentation. Recently, texture characterization based on the discrete wavelet transform (DWT) has attracted much attention and was found useful for a variety of texture-related applications, including classification, segmentation, and synthesis. Wavelet-domain Hidden Markov models (HMMs), in particular, the hidden Markov tree (HMT), have been recently proposed in [5] and applied to multiscale segmentation in [3]. The HMT can effectively characterize the joint statistics of wavelet coefficients by capturing interscale dependencies via tree-structured Markov chains across DWT scales. In [6], an improved wavelet-domain HMM, HMT-3S, was developed to capture dependencies both across subbands and across scales. In the HMT-3S, the three DWT subbands are integrated into one tree structure with the help of graphical grouping technique. It was shown that the more complete DWT characterization from the HMT-3S improves the accuracy of texture classification. Here we show that the HMT-3S can also improve the performance of multiscale Bayesian segmentation by providing more accurate statistical texture characterization.

This paper is organized as follows. Firstly, we briefly review multiscale Bayesian segmentation. Then we discuss the JMCMS algorithm and wavelet-domain HMMs. At last, the experiment results on two synthetic mosaics show that both contextual modeling and texture characterization play important roles in context-based multiscale Bayesian segmentation.

2. MULTISCALE BAYESIAN SEGMENTATION

We briefly review the multiscale segmentation approaches in [1, 2, 3] as follows. Given a random field \( Y \), we need to accurately estimate the pixel label in \( X \) where each label specifies one of \( N \) possible classes. With certain assumptions of prior distributions, Bayesian estimators attempt to minimize the average cost of an erroneous segmentation, as shown in the following,

\[
\hat{x} = \arg \max_x E[C(X, x)|Y = y]
\] (1)

where \( C(X, x) \) is the cost of estimating the true segmentation, \( X \). The MAP estimate is the solution to (1), if we use the cost function of \( C_{MAP}(X, x) = 1 \) whenever any pixel is incorrectly classified. It is known that the MAP estimator is excessively conservative. Therefore, multiscale Bayesian segmentation was proposed in [1], where an alternative cost function, called sequential MAP (SMAP) cost function, was introduced by proportionally summing up the segmentation errors from multiple scales together. The SMAP estimator aims at minimizing the spatial size of errors, resulting in more realistic results with lower computational complexity than the MAP estimator. The multiscale image model proposed in [1] is composed of a series of random fields at multiple scales. Each scale has a random field of image feature vectors, \( Y^{(n)} \), and a random field of class labels, \( X^{(n)} \). We denote an individual sample at scale \( n \) by \( y^{(n)} \) and \( x^{(n)} \), where \( s \) is the position in a 2-D lattice \( S^{(n)} \). By assuming Markovian dependencies across scales, the SMAP recursion can be computed in the fashion of coarse-to-fine as,

\[
\hat{x}^{(n)} = \arg \max_{x^{(n)}} \{ \log p_{y^{(n)}} x^{(n)} [y] + \log p_{x^{(n)} x^{(n+1)}} (x^{(n)} | x^{(n+1)}) \}.
\] (2)
The two terms in (2) are the likelihood function of image feature $y^{(n)}$ and the contextual prior knowledge from the next coarser scale, respectively. Specifically, the quadtree pyramid was developed in [1] to capture interscale dependencies of multiscalar class labels regarding the latter part of (2). Based on the same framework, a trainable context model for multiscale Bayesian segmentation was proposed in [2], where the contextual behavior can be trained off-line by providing image data and corresponding ground truth segmentations. Then the segmentation can be accomplished efficiently via a single fine-to-coarse-to-fine iteration through the algorithm was proposed in [3] where the context model is characterized by a context vector $v^{(n)}$, which means a dyadic image block in scale $\frac{\log_2 n}{2}$, as shown in Fig. 1(c). For an image decomposed from a set of neighboring samples $(3 \times 3)$ in the coarsest scale. It is assumed that, given $y^{(n)}_s$, its context vector $v^{(n)}_s = \{x^{(n)}_s, x^{(n)}_s\}$ can provide supplementary information regarding $x^{(n)}_s$, where $x^{(n)}_s$ denotes the class label of the parent sample and $x^{(n)}_s$ is dominant class label of the $3 \times 3$ samples at the coarsest scale. So given $y^{(n)}_s, x^{(n)}_s$ is independent with all other class labels. In particular, the contextual prior $P(x^{(n)}_s|y^{(n)}_s)(\text{du})$ is involved in the SMAP estimation which has the same purpose as the latter term in (2), and it is estimated by maximizing the following context-based mixture model likelihood as,

$$f(y^{(n)}|x^{(n)} = u) = \prod_{s \in \gamma^{(n)}} \sum_{c=1}^{N_c} p_{x^{(n)}_s, y^{(n)}_s} (c | x^{(n)}_s = u) f(y^{(n)}_s|x^{(n)}_s = c),$$

where the likelihood function $f(y^{(n)}|x^{(n)} = c)$ is computed using the wavelet-domain hidden Markov tree (HMT) proposed in [5]. An iterative Expectation Maximization (EM) training algorithm was developed in [3] to approach the above problem. The HMT is a tree-structured hidden Markov model in the wavelet-domain to characterize the joint statistics of wavelet coefficients across scales, as shown in Fig. 1(a). Correspondingly, Fig. 1(b) is a quadtree-structured wavelet subtree characterized by the HMT. In order to perform multiscale segmentation, an image is recursively divided into four sub-images of same size $J$ times and represented in a pyramid of $J$ scales, as shown in Fig. 1(c). For an image $I$ decomposed from a $J$-scale Haar DWT, we denote a block as $y^{(n)}_s$ which means a dyadic image block in scale $n$ of the pyramid representation. Given a set of Haar wavelet coefficients $\mathbf{w}$ of a test image and a set of HMT model parameters $\theta = \{\theta^{LH}, \theta^{HL}, \theta^{HH}\}$ of three wavelet subbands, i.e., $LH, HL$, and $HH$, the block $y^{(n)}_s$ is associated with three wavelet subtrees in three subbands as $\{T^{LH}_s, T^{HL}_s, T^{HH}_s\}$. The computation of the model likelihood of $f(y^{(n)}|\theta)$ is a realization of the HMT model $\theta$ and is obtained by

$$f(y^{(n)}|\theta) = f(T^{LH}_s|\theta^{LH}) f(T^{HL}_s|\theta^{HL}) f(T^{HH}_s|\theta^{HH}),$$

where it is assumed that three DWT subbands are independent and each one in (4) can be computed based the closed formula in [5]. Similarly, for all blocks in the pyramid, we can compute their model likelihood functions which can be used as the multiscale texture characterization in (3).

### 3. Joint Multi-Context and Multiscale

The simulation results in [1, 2, 3] show that the segmentation performance in homogeneous regions are usually better than those around texture boundaries. This is mainly owing to the fact that the context models used in those approaches mainly capture interscale dependencies and encourage the formation of large uniformly classified regions with less consideration on texture boundaries. To improve the segmentation results around texture boundaries, a joint multi-context and multiscale (JMCMS) approach was developed in [4], where multiple context models of different advantages are adopted. In the following, we discuss the context models of different structures with respect to their segmentation performance.

Given a sample $x^{(n)}_s$, its contextual information may come from some “neighbors” in the spatial and/or scale spaces. Then we naturally have three non-overlapped contextual sources as $P = x^{(n)}_s, NP = x^{(n)}_s$, and $N = x^{(n)}_s$, where $bs$ is the $3 \times 3$ window centered at $gs$ and excluding $gs$ at scale $n + 1$, and $hs$ is the $3 \times 3$ window centered at $s$ and excluding $s$ at scale $n$. Specifically, $P$ is the class label of $gs$, and $PN$ and $N$ are dominant class labels of $bs$ and $hs$, respectively. Based on $P, PN$, and $N$, we can have five low-complexity context models as shown in Fig. 2. It was found in [4] that the five context models have different segmentation results in terms of the accuracy of classification $P_a$, the accuracy of boundary localization $P_b$, and the accuracy of boundary detection $P_c$. A good segmentation requires high $P_a$, $P_b$, and $P_c$. Even though $P_a$ is usually the most important one, high $P_b$ and $P_c$ provide more desirable segmentation results with high accuracy of boundary localization and detection. From the extensive simulation results on a set of synthetic mosaics in [4], it was found that none of five context models can work well singly in terms of the three criteria, and they have different strengths in terms of the three criteria. For example, Context-2 has the best $P_a$, but the weakest in $P_b$. Context-5 is the strongest in $P_b$ but the weakest in $P_a$. Context-3 gives the highest $P_b$, but $P_a$ and $P_c$ suffer. Hereby a natural idea is to apply multiple context models together to obtain high $P_a$, $P_b$, and $P_c$.

Generally speaking, given $y = \{y^{(n)}|n = 1, 2, \ldots, L\}$ the collection of multiscale random fields of an image $Y$, a context model $V$ is used to simplify the characterization of the joint statistics of $y$ with the local contextual modeling. Thus, given $Z$ context models, we can have $Z$ different statistical characterizations of $y$, resulting in different segmentation results in terms of $P_a$. Fig. 1. (a) HMT (white node: the discrete hidden state variable $S$; black node: the continuous variable $W$. ) (b) Wavelet subtree. (c) Multiscale pyramid image representation.
model likelihood provides the multiscale texture characterization if we assume the hidden state number to be 2 for each wavelet coefficient, there are 8 states in a node of the HMT-3S. In other words, 2-state GMMS are still used to characterize the marginal DWT statistics in the 8-state HMT-3S. The EM training algorithm in [5] can be straightforwardly extended to the 8-state HMT-3S. The HMT-3S model likelihood can be computed as follows.

$$ f(\mathbf{w} | \theta_{HMT-3S}) = \sum_{k,i=0}^{N_k-1} \log \left( \sum_{u=0}^{7} f_u(T_{j,k,i} | \theta_{HMT-3S}, u) \right), $$

where $T_{j,k,i}$ is the complex wavelet subtree rooted at $w_{j,k,i} = \{w_{j,k,i}^{H}, w_{j,k,i}^{L}, w_{j,k,i}^{HL}, w_{j,k,i}^{LH} \}$, as shown in Fig. 3, and $f(T_{j,k,i} | \theta_{HMT-3S}, u)$ can be computed in a recursive fine-to-coarse fashion as follows,

$$ f_u(T_{j,k,i} | \theta_{HMT-3S}, u) = p_j(u)g(w_{j,k,i} | u) \left( \prod_{k=2i} f_u(T_{j-1,k,i} | \theta_{HMT}, v) \right), $$

and in the finest scale, i.e., $j = 1$, we have

$$ f_u(T_{1,k,i} | \theta_{HMT-3S}, v) = p_1(v)g(w_{1,k,i} | v), $$

and

$$ g(w_{j,k,i} | v) = \prod_{i \in B} g(w_{j,k,i}^{H} | 0, \sigma_{H,j,k}^{2}), $$

where $b = S_H \& v$, with $S_H = 1$, $S_L = 2$, and $S_H = 4$. In [6], it was shown the HMT-3S improves the accuracy of texture classification by capturing more complete cross-correlation of DWT. Similar to [3], the likelihood defined (6) can also be used in (2) and (3) to implement Bayesian segmentation.

5. SIMULATION RESULTS

The JMCMS (Context-2-3-5) and the HMT-3S are tested on two synthetic mosaics shown in Fig. 4. For comparison, we also study the segmentation algorithm in [3], namely HMTseg, where one Context-2 and the HMT model are used. We fix total iteration numbers of HMTseg and JMCMS to be the same, e.g., 30. Thus they have the similar computational complexity. We show the segmentation results of two mosaics in Fig. 4, and the numerical results in terms of $P_{\text{e}}$, $P_{\text{p}}$, and $P_{\text{c}}$ are given in Table 1. It is shown that both the JMCMS and the HMT-3S can improve the segmentation results over the HMTseg in terms of $P_{\text{e}}$, $P_{\text{p}}$, and $P_{\text{c}}$ by emphasizing the two terms in (2), respectively. Moreover, the integration of the JMCMS and the HMT-3S provides the best results regarding $P_{\text{e}}$ and $P_{\text{c}}$ which are considered the most important criteria.

As shown above, context-based multiscale Bayesian segmentation can be strengthened by two approaches. One is to use more sophisticated contextual modeling to characterize dependencies of multiscale class labels in terms of the latter part of (2). The other
Table 1. Segmentation performance comparison (%).

<table>
<thead>
<tr>
<th>Texture Models</th>
<th>Context Models</th>
<th>HMTseg</th>
<th>JMCMS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HMT</td>
<td>HMT-3S</td>
<td>HMT</td>
</tr>
<tr>
<td>Mosaic-A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$P_0$</td>
<td>88.10</td>
<td>91.53</td>
<td>95.08</td>
</tr>
<tr>
<td>$P_1$</td>
<td>11.47</td>
<td>17.79</td>
<td>31.41</td>
</tr>
<tr>
<td>$P_2$</td>
<td>51.09</td>
<td>49.26</td>
<td>53.77</td>
</tr>
<tr>
<td>Mosaic-B</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$P_0$</td>
<td>69.84</td>
<td>80.62</td>
<td>73.37</td>
</tr>
<tr>
<td>$P_1$</td>
<td>19.51</td>
<td>27.00</td>
<td>34.37</td>
</tr>
<tr>
<td>$P_2$</td>
<td><strong>65.90</strong></td>
<td>56.84</td>
<td>50.96</td>
</tr>
</tbody>
</table>

Fig. 4. Segmentation results of JMCMS and HMT-3S.

is to use more accurate texture features or texture model in terms of the former part of (2).

Basically, the methods in [1, 3, 4] mainly concern the supervised segmentation case where all texture features are given prior to segmentation regarding $\theta$ in (4) or (6). On the other hand, in [8], we also discussed the cases of semi-supervised and unsupervised segmentation where only partial or no texture features are given. Both HMT and JMCMS are involved for the SMAP estimation in [8]. Particularly, the K-mean clustering algorithm was used to identify training samples for the unknown textures from the clustering results of HMT model likelihoods. Then the self-supervised method is adopted to estimate the texture features regarding their HMT model parameters based on which the supervised segmentation process can be implemented. It was shown that the accurate segmentation results can be obtained for both semi-supervised and unsupervised segmentation scenarios.

6. CONCLUSIONS

This paper has discussed context-based Bayesian segmentation in terms of multiscale contextual modeling and statistical texture characterization, particularly, the JMCMS approach and wavelet-domain HMT-3S. The JMCMS is able to accumulate contextual behavior both across scales and via multiple context models, allowing more robust Bayesian estimation than a single context model. Compared with the HMT, the HMT-3S provides more accurate statistical texture characterization by capturing statistical dependencies both across subbands and across scales. We conclude that both multiscale contextual modeling and statistical texture characterization play important roles in multiscale Bayesian segmentation.

7. REFERENCES