SAR Image Labeling with Hierarchical Markov Aspect Models

Dengxin Dai, Wen Yang, Bill Triggs

***Abstract*—Scene segmentation and semantic labeling are important problems in SAR image interpretation. This paper proposes an efficient SAR imagery labeling method based on aspect model which can be learnt from keywords-labeled training data directly. Furthermore, a novel hierarchical Markov aspect model (HMAM) is presented by building aspect model on quadtree. HMAM outperform both aspect model and hierarchical MRFs due to their complementary as aspect model use global relevance estimates while quadtree can further explore image context and multi-scale cues. The experimental results on TerraSAR-X dataset show that our labeling method is effective and efficient, and demonstrate that HMAM improve labeling performance significantly with only a modest increase in learning and inference complexity than aspect model.**

***Index Terms*—Synthetic Aperture radar(SAR),Image labeling, Hierarchical markov aspect model,Quadtree.**

I. INTRODUCTION

Over the last decade we have witnessed an explosion in the number and throughput of airborne and spaceborne SAR imaging sensors. At the same time, advance in data transmission and store have made it increasingly possible to acquire and order SAR image data at a lower cost. The evergrowing large volumes of SAR images place a heavy demand on providing an effective and efficient scene segmentation and semantic labeling algorithm for understanding SAR imagery, which is aimed at partitioning a SAR image into their constituent semantic-level regions and assign appropriate class label to each region.This task is challenging because of the well-known “aperture problem”of local ambiguity [1]. For example, a homogenous dark region in SAR image maybe a piece of calm water, radar shadow, or road surface. Fig.1 shows two similarly dark regions. These two patches are not easily distinguishable without using context cues. Actually, one of them is the radar shadow owning to obscuration by the buildings within the illuminating radar beam, the other is water surface.To resolve this ambiguity, it is necessary to look at the patch within the context of a larger image area surrounding it.Therefore, we need to integrate multi-scale context information to overcome this local semantic ambiguity.

Though considerable progresses have been made in natural image labeling recently [2][3][4], SAR imagery labeling is still in a relative low-level stage. Compared with natural image labeling, SAR imagery labeling has several additional difficulties. Firstly, SAR imagery suffers from a noise-like phenomenon known as speckle which imposes a significant limitation on the measurement accuracy of resolution cell. Therefore, the extraction of robust appearance features for SAR Imagery is more difficult. Secondly, large-scale of SAR image pose a higher efficiency requirement to labeling algorithm. The last but not the least, researchers unable to validate their ideas adequately because there is no a publicly available SAR imagery interpretation dataset and it is tasteless and time-consuming to manually construct one’s own dataset.

Statistical distribution models with the maximum likelihood classification methods are well known and widely used in SAR image segmentation and classification. However, they are pixel-based methods which cannot handle the abundant information of SAR imagery and usually produce a characteristic and inconsistent salt-and-pepper labeling map.Many works built their models on MRFs to involve spatial relationship, often with remarkable improvement. Venkatachalam et al.[5] applied the wavelet-domain Hidden Markov Tree (HMT) models as a reliable initial segmentation, and then refined the classification using a multiscale fusion technique. Deng et al.[6] used a function-based parameter to weight the two components in a MRFs model and produced accurate unsupervised segmentation results for SAR sea ice images. Yang et al. [7] proposed a region-determined hierarchical MRF model for SAR image classification based on watershed over-segmentation algorithm, and demonstrated better results than the pixel-based hierarchical MRFs model. Xia et al.[8] presented a precise segmentation of SAR images using Markov random field model on region adjacency graph (MRF-RAG), and a rapid clustering method for SAR image segmentation was further proposed in [9], which embedded a Markov random field (MRF) model in the clustering space and used graph cuts for optimization. Li et al.[10] proposed to segment the SAR image based on the wavelet mixture heavy-tailed model and the hidden-class-label MRFs. These work only tested their methods on small images and had not yet listed their quantitative results. Our goal is to design an effective and efficient labeling algorithm which can handle the large-scale SAR images.

Aspect model such as PLSA [11] and LDA [12] are statistical models that are well appropriate to this task which can capture thematic coherence (image-wide correlations) and can resolve some cases of visual “polysemy ”. They posit each image is a mixture of a small number of latent topics and the creation of each word (patch) in image is attributable to one of the topics. Each image has its own mixing proportions but the topics and their attributes are shared by all images. We

---

D.X.Dai and W. Yang are with the Signal Processing Lab, School of Electronics Information, Wuhan University, Wuhan, 430079 China e-mail: ddx2004@gmail.com;yangwen94111@yahoo.com.cn.

B.Triggs is with Laboratoire Jean Kuntzmann,B.P. 53, 38041 Grenoble Cedex 9, France e-mail: Bill.Triggs@imag.fr

Manuscript received Jun. ?, ?; revised Dec. ?, 20??.

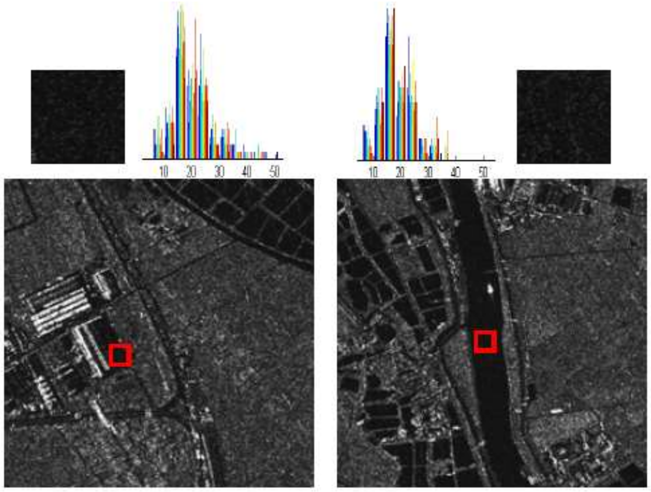


Fig. 1. Top: two ambiguous image patches and their histograms. Bottom: two images that contain the patches. Multi-scale cues and image context are helpful for labeling the patches.

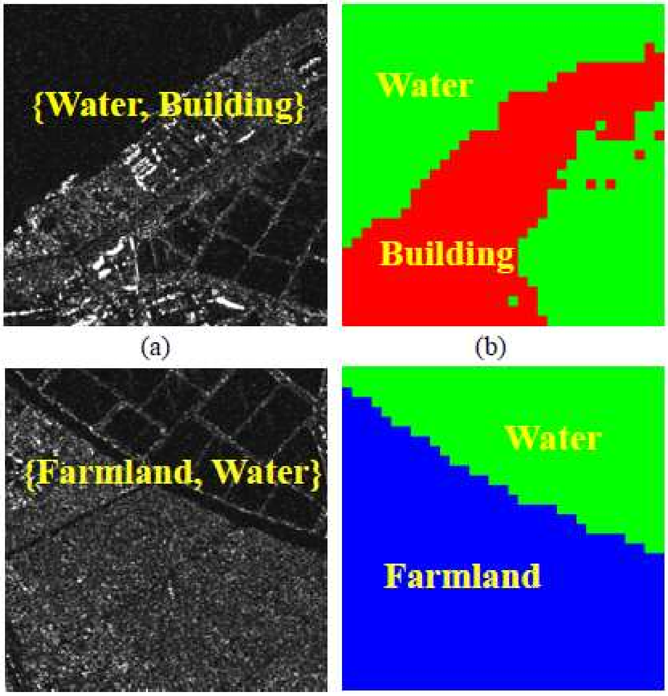


Fig. 2. Two keywords-labeled training images and the corresponding region labelings inferred during learning.

propose an efficient SAR imagery labeling method based on aspect model which can learn from keywords-labeled training images (as shown in Fig. 2) directly.

However, as a model proposed for document analysis, aspect model has its own limitation on image labeling. Firstly, there are not apparent visual words in image. Researchers often obtain visual words by clustering the features extracted from image patches at a single scale which leads to failure in capturing the instinct multi-scale cues in image. Moreover, aspect models assume that the labels of adjacent patches are independent, thus ignoring the strong correlations that are found in real image. Verbeek and Triggs [13] developed two spatial extensions of PLSA, but multi-scale cues were not considered. We develop a further extension, HMAM, based on quadtree which can explore multi-scale cues, spatial coherence and thematic coherence simultaneously. Our experimental results show that HMAM outperform PLSA significantly in SAR image labeling.

Our contributions are as follows. Firstly, an efficient SAR imagery labeling method is proposed which avoids the labor-intensive and time-consuming work to label every pixel in SAR imagey for obtaining detailed pixel-level training data. To our best knowledge, this is the first work that performs SAR imagery labeling on keywords-labeled training data. Secondly, we propose a hierarchical Markov aspect model (HMAM) on quadtree and apply it to SAR imagery labeling; Thirdly, we build a TerraSAR-X imagery dataset with which practitioners can evaluate their labeling algorithms quantitatively.

The rest of the paper is organized as follows. Section II reviews aspect models and describes our labeling methods based on aspect model. Section III illustrates Markov image modeling on quadtree and demonstrates hierarchical Markov aspect model. Section IV is devoted to our dataset and experimental results, and we draw our conclusion in section V.

II. ASPECT MODEL AND IMAGE LABELING

In this section, firstly, aspect model, the foundation on which our algorithm is built, is reviewed in some detail. Then, SAR image labeling method based on aspect model is described.

A. Aspect Model

Aspect model such as PLSA and LDA are statistical tools designed to analyze nature language from document collection $D = \{d_1, \ldots, d_N\}$ [11]. PLSA is a popular generative model which introduces a set of latent variables $z_k \in \{z_1, \ldots, z_K\}$ to explain data generation process. Each document d_i owns a specific mixing weight $P(z_k|d_i)$ over latent variables and each latent variable has a particular distribution $P(w_j|z_k)$ over the V words of dictionary. V is the total number of clusters (words) obtained by clustering image features. Here, document is represented as a bag of words sampled from a document-specific mixture of aspect model distribution:

$$P(w_j|d_i) = \sum_{k=1}^{K} P(w_j|z_k)P(z_k|d_i) \tag{1}$$

As a result the occurrence probability of document collection D is:

$$P(D) = \sum_{i=1}^{N}\sum_{j=1}^{V} P(d_i) \sum_{k=1}^{K} P(w_j|z_k)P(z_k|d_i) \tag{2}$$

where $P(d_i)$ is used to denote the probability that a word occurrence will be observed in document d_i. LDA is the Bayesian form of PLSA by adding a Dirichlet prior to the mixing weights, which makes a significant improvement for small documents but with heavy computational cost[13][14]. In this paper, we adopt PLSA for computational efficiency. The learning and inference for PLSA model can be completed by EM algorithm [15]. For the limitation of space, readers are referred to [11] for details.

B. Image Labeling based on Aspect Model

Generally, it is difficult to process a whole scene SAR imagery due to its huge size. Here, we treat a SAR imagery as a document collection D by partitioning it into hundreds of subimages and regarding each subimage as a document d. From each subimage we extract non-overlapping patches on a grid, representing them by feature descriptors. Visual words are obtained through clustering the extracted image features. Label inference performed at the patch level, but the results are propagated to pixel level for visualization and performance evaluation. We treat topics as scene categories (e.g. building, water). Labeling procedure is to be divided into two stages: training and inference. In training stage, the topic specific distribution $P(w|z)$ can be counted directly from pixel-level training data if they are available. It also can be learnt from the set of keywords-labeled training images shown as Fig.2. In this situation, we learn $P(w|z)$ from keywords-labeled images simply by setting $P(z_k|d_i)$ to zero for class k that is not emerged in the keywords list of document d_i. So only the images that are labeled with a topic contribute to learning its topic vector. The remaining $P(z_k|d_i)$ has a non-negative and sum-to-one value. Section IV shows even such weak supervision allows good topic models to be learnt. In inference stage, the topic specific distribution $P(z|w, d^{test})$ are computed. This is achieved by running EM in a similar way to learning stage, but now only the coefficients $P(z|d^{test})$ are updated in each M-Step with the learnt $P(w|z)$ kept fixed. These $P(z|w, d^{test})$are then used to label test images by likelihood maximization. The pipeline of the labeling method is illustrated in Fig.3. We address the labeling method based on aspect model learnt from pixel-level training data as PAM and state the method learnt from keywords-labeled training data as KAM.

III. HIERARCHICAL MARKOV ASPECT MODEL

Image patches often cause ambiguity when only based on local information. Fortunately, multi-scale cues and image context can make it clear what these patches are. Therefore, image labeling requires information coming from different scales and contextual information. Aspect model ignore the spatial structure of the image, modeling its patches independently at a single scale. In this section, we first discuss Markov image modeling and inference on patch-based quadtree. Then, the definition of HMAM on patch-based quadtree is described in detail.

A. Markov Image Modelling on patch-based Quadtree

To make it is feasible to define aspect model at multi-scale, we employ quadtree image representation proposed in [16] with some modifications. First, our finest resolution cell is a pre-selected patch of $S \times S$ pixels; Second, we only adopt several levels of quadtree for image modeling. We address this modified quadtree as patch-based quadtree which is illustrated in Fig.4. Now, we introduce how we model an image on this quadtree. The observed data Y are a multiresolution representation of the observed image, where the finest resolution on quadtree is consist of non-overlapping patches of $S \times S$ pixels partitioned from original image and the coarser-resolution is consist of patches with double size on each side. Each node on quadtree represent a patch in image. Thus, Y is a stochastic process indexed by the nodes of a quadtree as shown in Fig.4, where the set of nodes in the quadtrees is denoted B. The class label quadtree X is defined on the same multiresolution lattice as Y. Each level in the quadtree corresponds to a different spatial resolution level, where the top level represents the coarsest level 0, and level L the finest level. As usual in statistical approaches, x and y are viewed as occurrences of random fields X and Y, where Y is the space of observe states and X is the space of class labels. As shown in Fig.4, any nodes b except those at the coarsest level has a unique parent node b^-. Conversely, the set of the four children of any node b is denoted b^+. The set of descendants of b, including b itself, is denoted $d(b)$.

The inference of X is performed by an extension of Viterbi algorithm [17]. This algorithm is noniterative and requires two passes on the tree. Now, we deduce the posterior marginal $P(x_b|y)$ for our patch-based quadtree. We yield the expression of the patch posterior marginal $P(x_b|y)$ as a function of the posterior marginal at parent node b^-:

$$\begin{aligned} P(x_b|y) &= \sum_{x_{b^-}} P(x_b|x_{b^-}, y)P(x_{b^-}|y) \\ &= \sum_{x_{b^-}} P(x_b|x_{b^-}, y_{d(b)})P(x_{b^-}|y) \qquad (3) \\ &= \sum_{x_{b^-}} \frac{P(x_b, x_{b^-}|y_{d(b)})}{\sum_{x_{b^-}} P(x_b, x_{b^-}|y_{d(b)})} P(x_{b^-}|y) \end{aligned}$$

This yields a top-down recursion provided that the posterior marginal $P(x_0|y)$ at the coarsest level, as well as probabilities $P(x_b, x_{b^-}|y_{d(b)})$ are made available. Because of the noncausal structure at the coarsest level of the patch-based model, we obtain posterior probability $P(x_0|y)$ using LBP algorithm [19] at this level. Another part is obtained as :

$$P(x_b, x_{b^-}|y_{d(b)}) = P(x_{b^-}|x_b)P(x_b|y_{d(b)}) \qquad (4)$$

where the first factor is derived as:

$$P(x_{b^-}|x_b) = P(x_b|x_{b^-})P(x_{b^-})/P(x_b) \qquad (5)$$

where $P(x_b|x_{b^-})$ is the pre-defined transaction probability, and $P(x_b)$ is computed by a simple top-down recursion:

$$P(x_b) = \sum_{x_{b^-}} P(x_b|x_{b^-})P(x_{b^-}) \qquad (6)$$

$P(x_b|y_{d(b)})$in formula (4) is computed through a bottom-up procedure:

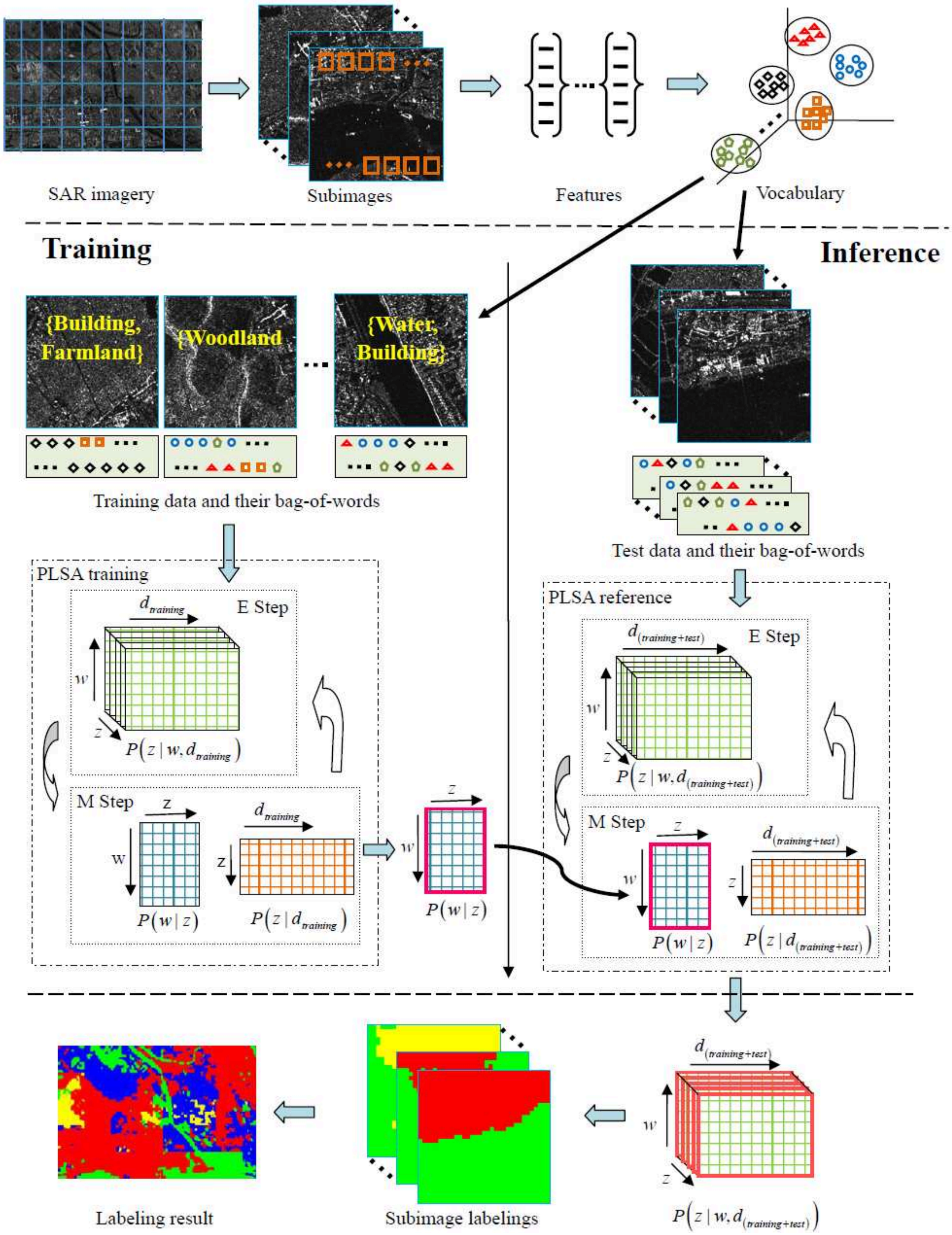


Fig. 3. Pipeline of SAR imagery labeling method based on aspect model.

$$
\begin{aligned}
&P(x_b|y_{d(b)}) \propto P(x_b, y_{d(b)}) \\
&= \sum_{x_{b+}} P(y_{d(b)}|x_b, x_{b+}) P(x_{b+}|x_b) P(x_b) \\
&= \sum_{x_{b+}} P(x_b) \prod_{t \in b^+} \left[P(y_{d(t)|x_t}) P(x_t|x_b) \right] \qquad (7) \\
&\propto P(y_b|x_b) P(x_b) \prod_{t \in b^+} \sum_{x_t} \left[\frac{P(x_t|y_{d(t)})}{P(x_t)} P(x_t|x_b) \right]
\end{aligned}
$$

B. Hierarchical Markov Aspect Model

Inspired by the work of [18],we propose modeling the topic (aspect) of visual words in image as a hidden Markov Tree. Specifically, we assume that neighboring patches are more likely to have the same topics, parent are more likely take the same topic with their children and vice versa. We build one patch-based quadtree with L level for each subimage d and extract features from the patches at each scale $l \in (0, 1, \ldots, L)$ separately. Every scale-specific feature descriptor is vector quantized into V bins using centers learnt by k-means from the same scale of all the quadtrees. The model fitting of HMAM consist of four steps:

1) Initialization: we run basic PLSA training/test procedure at every scale l independently until convergence, then, record them as $P^l(z|w, d)$.
2) Inference on quadtree: we use $P^l(z|w, d)$ to initialize $P^l(y_b|x_b, d)$ and then run quadtree inference. It is obvious that this substitution is reasonable. Here, z in aspect model and x in Markov modeling both indicate scene categories (e.g. building area, water area), and w and y both represent observed data in a particular subimage d. So, we have

$$
\begin{aligned}
P(z|w, d) &= P^l(x_b|y_b, d) \\
&\propto P^l(x_b, y_b, d) \qquad (8) \\
&\propto P^l(y_b|x_b, d)
\end{aligned}
$$

3) Maximization step: firstly, we deliver $P^l_{MPM}(x_b|y_b, d)$, the inference results of step (2), to $P^l_{MPM}(z|w, d)$ which has the same structure with $P^l(z|w, d)$. Then, we estimate $P^l(w|z)$ and $P^l(z|d)$ from $P^l_{MPM}(z|w, d)$ based on likelihood maximization, which are formulated as follows:

$$
P^l(w_j|z_k) = \frac{\sum_{i=1}^{N} n^l(d_i, w_j) P^l_{MPM}(z_k|w_j, d_i)}{\sum_{m=1}^{M} \sum_{i=1}^{N} n^l(d_i, w_m) P^l_{MPM}(z_k|w_m, d_i)} \qquad (9)
$$

$$
P^l(z_k|d_i) = \frac{\sum_{j=1}^{M} n^l(d_i, w_j) P^l_{MPM}(z_k|d_i, w_j)}{n^l(d_i)} \qquad (10)
$$

4) Expectation step: we simply apply Bayes formula and obtain,

$$
P^l(z_k|d_i, w_j) = \frac{P^l(w_j|z_k) P^l(z_k|d_i)}{\sum_{c=1}^{K} P^l(w_j|z_c) P^l(z_c|d_i)} \qquad (11)
$$

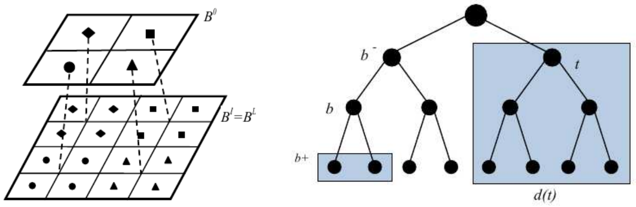


Fig. 4. Quadtree structure and notations on the tree.

Then, we check whether the algorithm is convergent. We terminate the recursion if it is true, and turn to setp 2), otherwise.

Top-down and bottom-up inference procedure in HMAM make knowing parent is farmland alter the conditional distribution of its children and vice versa. The posterior probability inferred at the coarsest level introduce image context explicitly and children who have the same parent also introduce image context implicitly. Image labeling method based on HMAM is similar to the labeling method described in section II. We can perform image labeling task in that framework with HMAM instead of aspect model. $P^L(z|w, d^{test})$ are adopted for visualization and performance quantification. We address the labeling method based on HMAM learnt from pixel-level training data as PHAM and take the method learnt from keywords-labeled training data as KHAM.

IV. EXPERIMENTAL RESULTS AND DISCUSSIONS

This section demonstrates our dataset and analyze the performance (accuracy and speed) of the proposed labeling methods on TerraSAR-X dataset in detail.

A. Datasets and Experimental Settings

Our experimental datasets are built on a whole scene TerraSAR-X image (48189×25255 pixels) of Foshan in central Guangdong province, China, acquired in 24/05/2008 (©Infoterra GmbH/DLR) at "stripmap"mode. The spatial resolution is about 3×3m. The pixel-level ground truth is labeled manually according to the corresponding optical remote sensing imagery(SPOT5) and related geographic information. Pixels are assigned to four classes: building, woodland, farmland and water. Pixels we are not sure which class they should belong to are labeled as "void ". About 13% of the pixels are unlabeled ("void") in the experimental images.We ignore void pixels during both training and evaluation. In our experiments, the whole imagery is partitioned into 1800 subimages (documents) of 960×960 pixels with 80 ($2^L \times S$) pixels overlapping. The overlapping pixels can maintain Markov property over the whole SAR imagery. These pixels are not taken into account in performance evaluation and final illustration of labeling results. We partition the dataset into 400 subimages for training and 1400 subimages for test, and report average results over 10 random train-test partitions. The patch is labeled to the class with which the posterior probability associated is the maximum and its ground truth is taken to be the most frequent pixel label within it. The labeling results are evaluated at pixel-level by linearly interpolating the 4 adjacent patch-level posteriors to pixels.

Four widely used features for SAR image segmentation, GLCM [20], Gabor filters [21], Gauss Markov random fields (GMRF) Texture [22] and histogram are employed in our experiments. The features used here are implemented with the following parameters. Histogram is used with 32 bins. For GLCM and GMRF we use the same parameters setting as [23]. Grey levels are quantized to 32, inter-pixel distance set to 1, and orientation set to 4. Four statistics are selected: Contrast, Entropy, Correlation, and Homogeneity. Gabor texture descriptors are used with 6 scales and 8 orientations based on an efficient implementation named "simple Gabor feature space"[24]. These four descriptors are quantized into 400 centers by K-means respectively. We set $S = 20$ to balance the tradeoff between robust representation and pixel-level labeling accuracy. Three pyramid levels are used (i.e., $L = 2$). More levels are tried, but with less further improvement. The transition probabilities are

$$\begin{cases} P(x_b = j | x_{b^-} = i) = \alpha & \text{if } i = j \\ P(x_b = j | x_{b^-} = i) = \frac{1-\alpha}{M-1} & \text{otherwise} \end{cases} \tag{12}$$

where M is the number of classes, and we set $\alpha = 0.9$. These parameters are set experimentally.

For performance evaluation of our labeling methods, two commonly used techniques, Maximum Likelihood(ML) and support vector machine (SVM) [25] classifiers are also adopted as benchmarks. As an implementation of SVM, we use the very public SVM package [26]. For fairly, the same train-test partitions and features are applied to SVM and it is performed with its optimal parameters selected experimentally. Gamma distribution has been widely used in SAR imagery modeling. Here, we take the conditional probability of pixel intensity for each class as a specific Gamma distribution which is used to label pixels by ML.

B. Qualitative results on TerraSAR-X dataset

Fig.5 demonstrates two labeling results (each 8800×6400 pixels) of KHAM with the corresponding groud truth on TerraSAR-X images, while Fig.6 presents four group labeling resutls on subimages using KAM and KHAM.The regions in Fig.5 are obtained both by merging 88 subimages (overlapping pixels are ignored), and we can find that our method has some mosiac effect which mainly because our patch-based representation. There is also some incorrect labeling on narrow river regions, combining with some river detection techniques may alleviate this deficiency.

TABLE.I shows a comprehensive comparison on classification accuracy using different classifier with different features. Accuracy values in the table are computed as percentage of image pixels assigned to the correct class label, ignoring pixels labeled as void in the ground truth. Here, we not list the performance of ML classifier with gamma distribution, which obtains the worst accuracy in this experiment, it is only 55.6%. It is mainly due to two reasons. The first is pixel scattering intensity cannot be characterized accurately due to speckle noise. Comparatively speaking, patch-based representation are more robust and informative, that is also why we focus our attention mainly on patch-based representation in our experiments. Another factor is that one semantic class may cover several types of homogenous regions which are different from each other explicitly in statistical characterization. The highest accuracy arrives at 85.3% using PHAM with histogram features.

TABLE I
LABELING ACCURACIES OF DIFFERENT LABELING METHODS WITH DIFFERENT FEATURES(%)

Method **Feature**	**SVM**	**PAM**	**KAM**	**PHAM**	**KHAM**
GLCM	64.5	70.3	69.1	**73.4**	71.9
Gabor	65.8	71.1	69.5	**74.9**	73.5
GMRF	69.1	74.8	74.5	**78.7**	77.6
Hist	70.3	82.7	81.0	**85.3**	84.1

C. Discussions

This section analyzes and discusses some details of our labeling method.

1) Benefits of Aspect Model: The first main conclusion from TABLE I is that our labeling methods are all significantly outperform SVM. With the same pixel-level training data, PAM exceeds SVM in labeling accuracy at least more than 5.3% no matter what features to use. Althourg used keywords-labeled training data, KAM also outperforms SVM by 3.7. This is mainly due to our methods take the advantages of aspect models which can capture thematic coherence (image-wide correlations) and can resolve some cases of visual polysemy. We may also benefit from bag-of-features (clustering) techniques which can discover image primitives and reduce noise effects at a certain degree.

2) Benefits of Incorporating Multi-scale Cues and Image Context: The second main conclusion from TABLE I is that HMAM is superior to aspect model at a single scale in image labeling. With the help of multi-scale cues and image context, PHAM increases labeling results by 3% for PAM, and KHAM also increases labeling results by 3% for KAM. TABLE II and III indicate that multi-scale information and image context are more important for building area, while provide less help for farmland, woodland and water regions. It is maily because different patches in farmland basically have the same statistical properties, and the same case in woodland and water area. Therefore, there is little complementary information between neighboring patches in these scenes, and also the same case between parent and children. However, patches in building area may different from each other significantly. Hence, some patches with high probabilities can disambiguate their neighbors, parent, or children. All in all, multi-scale cues and image context are more useful in dealing with complex scenes. Fig.6 illustrates four groups of subimage labelings of KAM and HKAM, from which we can see the incorporation of multi-scale cues and image context can really disambiguate some patches.

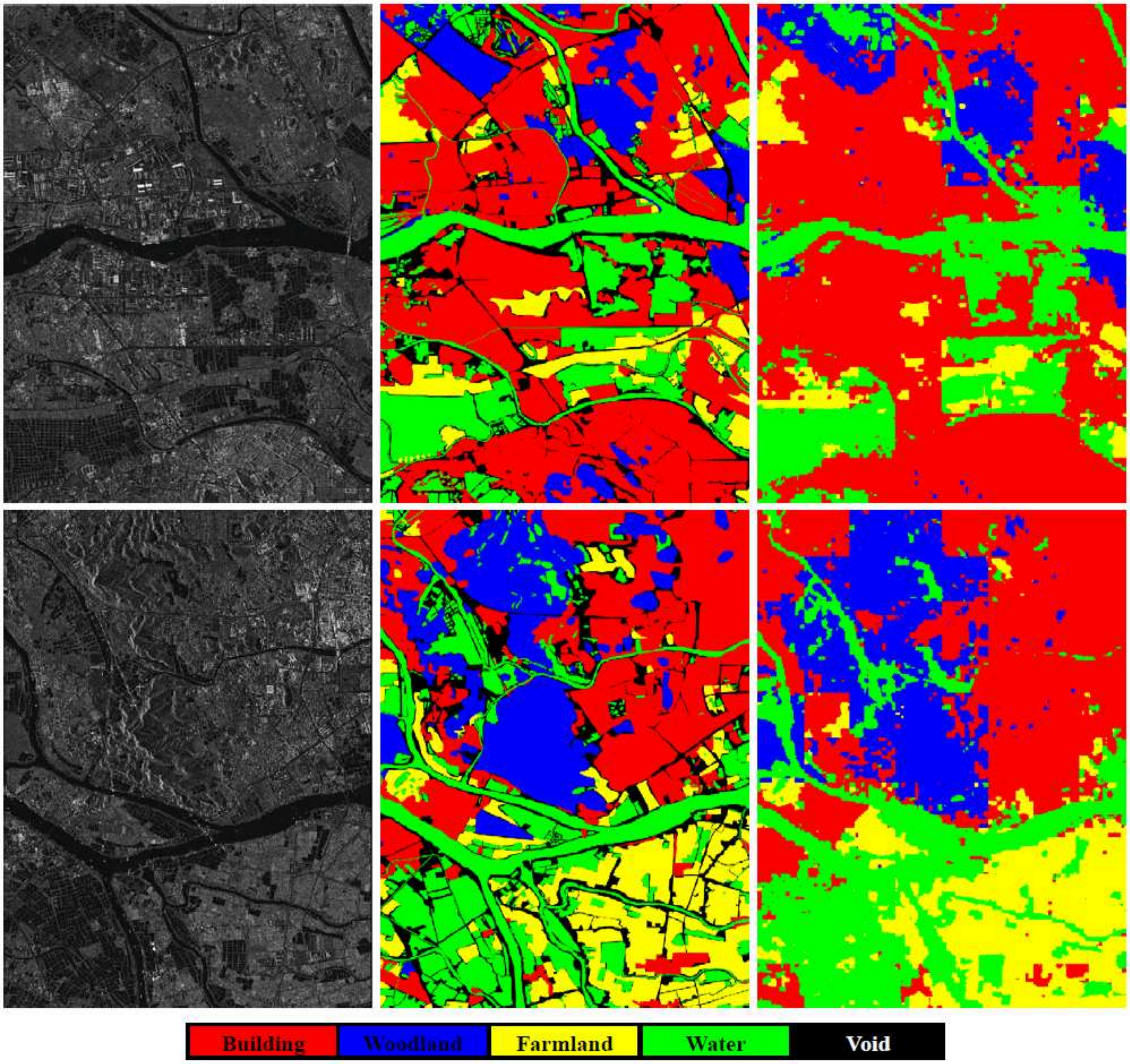


Fig. 5. TerraSAR-X image labeling results, the first column shows two regions (each 8800×6400 pixels), the second column illustrates the corresponding ground truth, the last column is our labeling results of KHAM.

TABLE II
LABELING RESULTS OF KAM(%)

Terrain Class	Building	Woodland	Farmland	Water
Building	**84.6**	6.5	4.4	4.5
Woodland	20.0	**63.7**	13.1	3.0
Farmland	8.0	6.6	**83.4**	2.0
Water	7.5	1.4	3.0	**88.2**

TABLE III
LABELING RESULTS OF KHAM(%)

Terrain Class	Building	Woodland	Farmland	Water
Building	**89.4**	3.3	3.9	3.3
Woodland	20.8	**64.8**	12.1	2.3
Farmland	8.2	4.7	**83.8**	3.4
Water	6.2	1.3	3.3	**89.3**

3) Benefits of learning from keywords-labeled data: The third main conclusion from TABLE I is that KAM and KHAM can achieve comparative performance to PAM and PHAM respectively, while the former only use keyword-labeled training data. It is a great merit for SAR imagery labeling, because it is expensive and labor-intensive to manually label each pixel in SAR images while it is convenient to obtain keywords-labeled training data. This property ensure the generalization of our methods to large-scale SAR imagery labeling.

4) Feature Selection and Speed: From TABLE I, we learn that histogram is a simple but informative descriptor for SAR imagery labeling. Texture-based features such as GLCM, Gabor and GMRF achieve lower performance than histogram in our experiments. Consequently, it is important to select the features that are most informative for separating land-cover

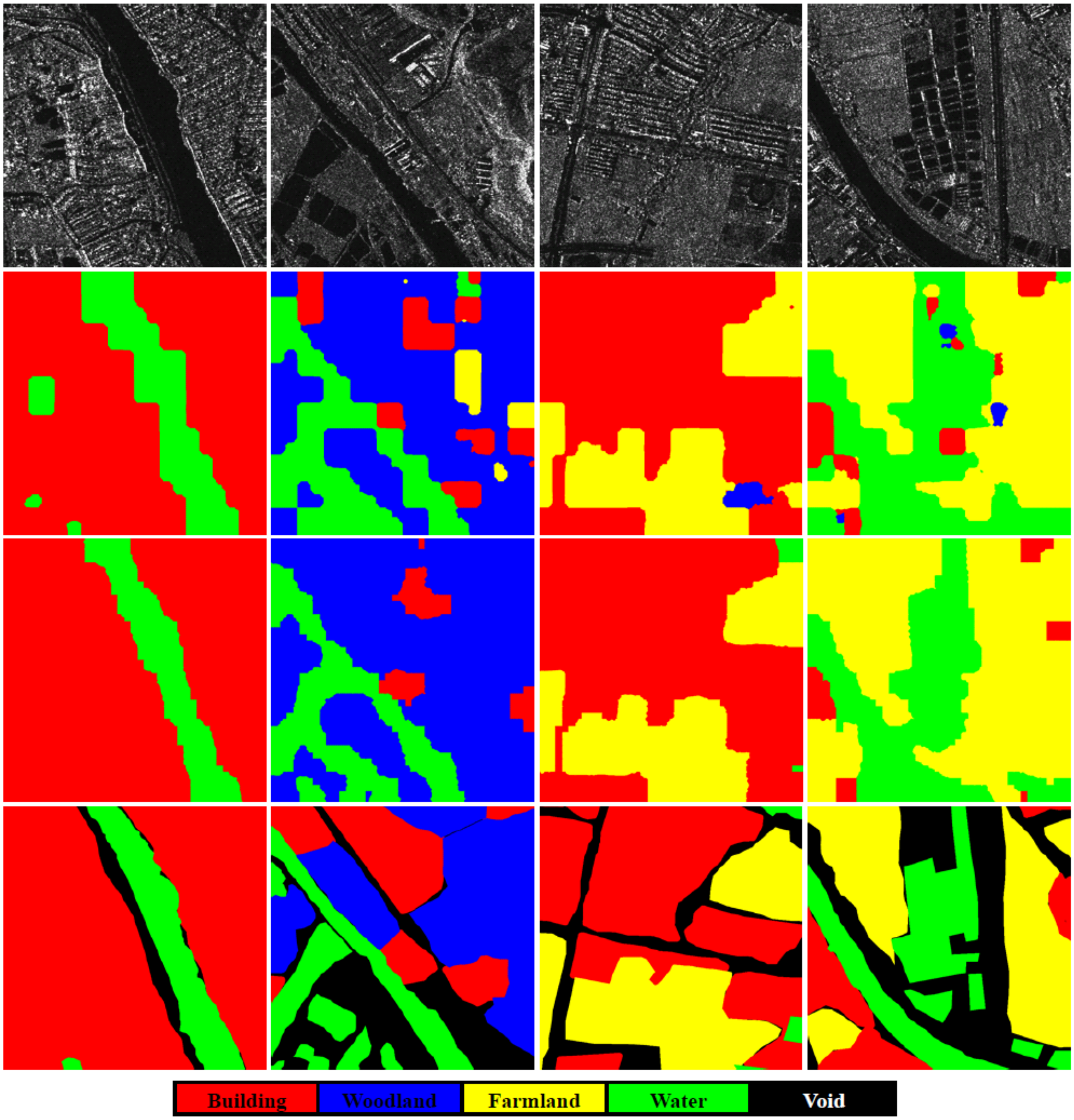


Fig. 6. Subimages labeling results with KAM and KHAM, the first row shows four subimges (800 × 800 pixels), the second and the third rows illustrate their labeling results with KAM and KHAM,respectively, and the final row shows their ground truth.

classes. We list the computing speed of our methods and the benchmarks in TABLE IV. Currently, our unoptimized matlab implementation runs on a 2.4-GHz Pentium-class machine with 4G memory. TABLE IV also conclude that our methods are efficient both in training and test. Compared to PAM and KAM, PHAM and KHAM require more training and test time. The increased expenses are directly proportional to L.

5) Labeling as function of the proportion of training data: We now consider how the performance of our labeling methods drop as the proportion of training data decreases. We varied the proportion of training data versus the whole data

TABLE IV
TRAINING AND TEST SPEEDS OF DIFFERENT LABELING METHODS

Method	Training Time	Test Time
ML	-	$<$0.01 sec/image
SVM	15 sec/image	$<$0.01 sec/image
PAM	-	$<$0.1 sec/image
KAM	$<$ 0.1 sec/image	$<$0.1 sec/image
PHAM	$<$0.05 sec/image	0.3 sec/image
KHAM	0.3 sec/image	0.3 sec/image

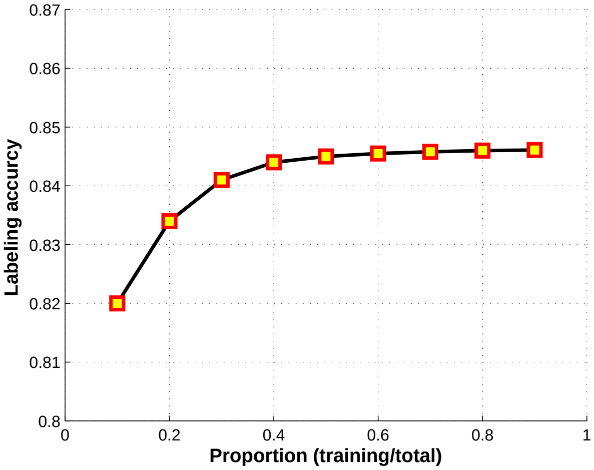


Fig. 7. Labeling accuracy of KHAM when learning from increased proportion of training data.

(training+test) from 0.1 to 0.9. PAM, KAM, PHAM, and KHAM have the very similar response. Here, for clarity, we only illustrate the experimental results of KHAM in Fig.7. We can conclude from Fig.7 that our labeling methods can achieve satisfactory performance even with small training dataset.

V. CONCLUSION

We addressed the challenge of labeling a whole SAR imagery, and proposed a solution by learning semantic classes from training images that are labeled with image-level keywords rather than with detailed pixel-level labeling. We showed that aspect models (modeling each image as a mixture of latent "aspect") are appropriate to SAR imagery labeling for its efficiency and effectiveness. We also extend aspect model to hierarchical Markov aspect model to incorporate multi-scale information of image. Experimental results on TerraSAR-X SAR image show that our labeling methods achieves impressive labeling performance, and validate our hierarchical extension of aspect model.

ACKNOWLEDGMENT

This work were supported by a grant from the National Natural Science Foundation of China (No.40801183,60890074) and the National High Technology Research and Development Program of China (No.2007AA12Z180,155)

REFERENCES

[1] S. Lazebnik and M. Raginsky, "An Empirical Bayes Approach to Contextual Region Classification,"*Proc. IEEE CS Conf. Computer Vision and Pattern Recognition*, to appear,2009.

[2] J. Shotton, J. Winnand, C. Rother, and A. Criminisi, "Textonboost: Joint Appearance, Shape and Context Modeling for Multi-Class Object Recognition and Segmentation,"in *Proc. Ninth European Conf. Computer Vision*, vol. 1, pp. 1-15, 2006.

[3] J.Verbeek, B. Triggs, "Scene Segmentation with CRFs Learned from Partially Labeled Images,"in *Advances in Neural Information Processing Systems*, pp.1553-1560, MIT Press, 2008.

[4] S. Gould, J. Rodgers, D. Cohen, G. Elidan, D. Koller, "Multi-Class Segmentation with Relative Location Prior,"*Int'l J. Computer Vision*, vol. 80, no.3, pp.300-316, 2008.

[5] V.VenkatachalamH.Choi, R.G.Baraniuk, "Multiscale SAR image segmentation using wavelet- domain hidden markov tree model,", in *Proceedings of the SPIE lJth International Symposium on Aerospace/Defense Sensing, Simulation, and Controls, Algorithms for Synthetic Aperture Radar Imagery VII*, vol.3497,pp.110-120,2000

[6] H.W. Deng, D.A.Clausi, "Unsupervised segmentation of synthetic aperture radar sea ice imagery using a novel Markov random field model,"*IEEE Trans. Geosci. Remote Sens.*, vol.43,no.3, pp.528-538,2005.

[7] Y. Yang, H. Sun, and C. He, "Supervised SAR image MPM segmentation based on region-based hierarchical model,"*IEEE Geosci. Remote Sens. Lett.*, vol.3,no.4, pp. 517-521, 2006.

[8] G.S.Xia,C.He,H.Sun, "Integration SAR Image Segmentation Method Using MRF on Region Adjacency Graph,", *IET Radar, Sonar and Navigation*, Vol.1, No.5, pp.348-354, 2007.

[9] G.S.Xia,C.He,H.Sun, "A Rapid and Automatic MRF-based Clustering Method for SAR Images,", *IEEE Geosci. Remote Sens. Lett.*, Vol.4, No.4, pp.596-600,2007.

[10] M. Li, Y. Wu, Q. Zhang, "SAR image segmentation based on mixture context and wavelet hidden-class-label Markov random field,"" *Computer & Mathematics with Applications*, vol. 57, no.6, pp. 961-969, 2009.

[11] T. Hofmann, "Unsupervised learning by probabilistic latent semantic anaylysis,"*Machine Learning*, vol.42,no.1-2, pp. 177-196, 2001.

[12] D. Blei, A. Ng, and M. Jordan, "Latent Direchlet allocation,"*Journal of Machine Learning Research*, vol. 3, no.3, pp.993-1022, 2003.

[13] J. Verbeek, B. Triggs, "Region Classification with Markov Field Aspect Models,"in *Proc. IEEE CS Conf. Computer Vision and Pattern Recognition*. 2007.

[14] F.-F. Li and P. Perona. "A bayesian hierarchical model for learning natural scene categories,"in *Proc. IEEE CS Conf. Computer Vision and Pattern Recognition*, pages 524C531, 2005.

[15] A. Dempster, N. Laird, and D. Rubin, "Maximum likelihood from incomplete data via the EM algorithm,"*Journal of the Royal Statistical Sociaty*, Series B, vol.39, pp. 1-38, 1977.

[16] J.M. Laferte, P. Perez, and F. Heitz, "Discrete Markov image modeling and inference on the quadtree,"*IEEE Trans. Image Process.*, vol.9, no. 3, pp. 390-404, 2000.

[17] G.D. Forney, "The Viterbi algorithm,"in *Proc. IEEE*, vol.61, no.3, pp. 268-278, 1973.

[18] A. Gruber, M. Rosen-Zvi and Y. Weiss,"Hidden Topic Markov Models,", in *Artificial Intelligence and Statistics (AISTATS)*, San Juan, Puerto Rico, March 2007.

[19] J. Yedidia, W. Freeman, and Y. Weiss, "Understanding belief propagation and its generalizations,"*Technical Report TR-2001-22*, Mitsubishi Electric Research Laboratories, 2001.

[20] R.M. Haralick, K. Shanmugan, and I. Dinstein, "Textural features for image classification,"*IEEE Trans. Syst. Man. Cybern*, vol.8,no.6, pp. 610-621, 1973.

[21] A.K. Jain, and F. Farrokhnia, "Unsupervised texture segmentation using Gabor filters,"*Pattern Recognition*. vol.24,no.12,pp. 1167-1186, 1991.

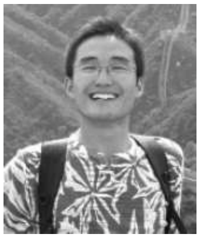
[22] R.Chellappa, "Two-dimensional discrete Gaussian Markov random field models for image processing,"in *Machine Intelligence and Pattern Recognition: Progress in Pattern Recognition 2*, G.T. Toussaint (Ed.), Elsevier Science Publishers, B.V. (North-Holland). pp.79C112,1985.

[23] D.A. Clausi, "Comparison and fusion of co-occurrence, Gabor, and MRF texture features for classification of SAR sea ice imagery,"*Atmosphere & Oceans*, vol.39, no.4, pp. 183-194, 2001.

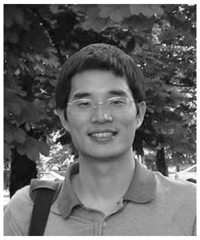
[24] V. Kyrki, J. K. Kämäräinen, H. Kälviäinen, "Simple Gabor Feature Space for Invariant Object Recognition,"*Pattern Recognit. Lett.*, vol. 25, no.3 pp. 311-318, 2004.

[25] C. Corinna and V. Vapnik, "Support-Vector Networks,"*Machine Learning*, vol.20, pp. 273-291, 1995.

[26] C.C. Chang and C.J. Lin, "LIBSVM : a library for support vector machines,"2001.Software available at:
http://www.csie.ntu.edu.tw/∼cjlin/libsvm



Dengxin Dai Dengxin DAI received B.S. degree in optical information science and technology in 2008, from Wuhan University, Wuhan, Hubei, China, where he is currently working toward the Ph.D degree at the signal processing laboratory, school of electronic information. His research interests include image processing, computer vision, and machine learning.



Wen YANG Wen YANG received his Ph.D. from Wuhan University, China, in 2004. He is currently an associate professor at the School of Electronic Information, Wuhan University. His research interests include image segmentation and classification, target detection and recognition, machine learning and data mining with applications to remote sensing.

Bill Triggs Biography text here.