A Multi-scale Latent Dirichlet Allocation Model for Object-oriented Clustering of VHR Panchromatic Satellite Images

Hong Tang*, Li Shen, Yinfeng Qi, Yunhao Chen, Yang Shu, Jing Li, David A. Clausi, *Member, IEEE*

***Abstract*—A novel model is presented to address the problem of semantic clustering of geo-objects in VHR panchromatic satellite images. The proposed model combines a probabilistic topic model with a multi-scale image representation into an automatic framework by embedding both document and scale selections. The probabilistic topic model is used to characterize the statistical distributions of both intra-class appearance and inter-class coherence of geo-objects within documents, i.e., squared sub-images. Because the bag-of-words assumption involved in the probabilistic topic models does not consider the spatial coherence between topic labels, the multi-scale image representation is designed to provide a self-adaptive spatial regularization for various geo-object categories. By introducing scale and document selections, the automatic framework integrates the probabilistic topic model and the multi-scale image representation to ensure that *words* on a site should be allocated the same topic label no matter what *documents* they reside in. Consequently, unlike the traditional method of applying topic models for analyzing satellite images, the process of explicitly generating a set of documents before modeling and then combining multiple labels for a word on a given site is unnecessary. Gibbs sampling is adopted for parameter estimation and image clustering. Extensive experimental evaluations are designed to first analyze the effect of parameters in the proposed model and then compare the results of our model with those of some state-of-the-art methods for three different types of images. The results indicate that the proposed algorithm consistently outperforms these exiting state-of-the-art methods in all of the experiments.**

***Index Terms*—Probabilistic topic models; object-oriented clustering; latent Dirichlet allocation; scale space theory.**

Manuscript received June 18, 2012. This work is supported by the National Basic Research Program of China (No. 2011CB707102), the National Natural Science of China (No. 40901217), the Program for New Century Excellent Talents in University (No. NECT-11-0039) and the Fundamental Research Funds for the Central Universities (No. 105563GK).

H. Tang, L. Shen, Y. Qi, Y. Chen, Y. Shu and J. Li are with the State Key Laboratory of Earth Surface Processes and Resource Ecology and Key Laboratory of Environment Change and Natural Disaster, Beijing Normal University, Beijing 100875, China (e-mail: hongtang@bnu.edu.cn; sandyshenli@gmail.com; xiaoqixyz@gmail.com; cyh@bnu.edu.cn; sysun@mail.bnu.edu.cn; lijing@bnu.edu.cn). D.A.Clausi is with the Department of Systems Design Engineering, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1 (e-mail: dclausi@uwaterloo.ca).

*Corresponding author: H. Tang. Tel: +86-10-58806401; Fax: +86-10-58806401.

I. Introduction

IMAGE clustering is a widely used technique for extracting geographic thematic information from remote sensing images when a training sample is not available [1-4]. A common statistical approach to the image clustering problem is to model the probability density function of individual pixel attributes (e.g., spectra, texture) using a finite mixture model (FMM) [5]. The clustering process is then alternatively converted to infer the most probable components (i.e., cluster labels) of the mixture model to generate pixels and estimate the parameters of the model. Although the statistical characteristics of pixel attributes might be well described in the mixture models, the cluster labels are spatially independent [6]. In other words, no spatial context information, e.g., local interactions or statistical characteristics among pixels, is considered in the models. Markov Random Field (MRF) models are often employed to embed a prior context constraint on the cluster labels to enhance the spatial consistency of the labels [7-14]. Instead of individual pixels, a set of squared sub-images are modeled using probabilistic topic models [15-18] to discover semantic patterns from Very High Resolution (VHR) satellite images in an unsupervised way [19, 20].

In this paper, we address the problem of object-oriented semantic clustering of geo-objects in VHR panchromatic satellite images in an automatic framework. The term "object-oriented" includes two meanings. First, object-oriented indicates that the proposed method differs from the pixel-based image analysis methods, which are based on processing the entire scene pixel by pixel. Second, object-oriented also indicates that the proposed method is based on squared images with a preset size, rather than segments of geographic geo-objects. In an object-based image analysis, the basic processing units are image segments that are obtained through image segmentation [17]. The proposed model consists of three complementary components: (1) a probabilistic topic model component that allows modeling of the statistical distributions of both the inter-object coherency and the intra-class appearance of geo-objects within documents, (2) a multi-scale image representation component that provides a self-adaptive

spatial regularization for various geo-object categories, and (3) an automatic application framework component that integrates the other two components using both document and scale selections. Although the probabilistic topic model has been employed to extract thematic information from remote sensing images in previous studies, the main contribution of the proposed model is the novel application framework of combining a probabilistic topic model with a multi-scale image representation to achieve effective clustering for VHR panchromatic satellite images.

The remainder of this paper is organized as follows. We first review related studies in section II. Then, we present the proposed model and algorithm for semantic clustering in section III. The experimental results and evaluation are presented in section IV, and we conclude with a discussion in section V.

II. Related Work

In this section, we first review probabilistic topic models in the text analysis community, and then describe their application for image analysis with a focus on both natural and satellite images.

Probabilistic topic models originated as a suite of language models from a seminal work [18], whose goal was to discover the latent patterns or structures (i.e., topics) of words from a large collection of documents for effective clustering or retrieval [21, 22]. Hoffman (1999) developed the first probabilistic topic model for analyzing latent topics from a set of documents, i.e., the probabilistic latent semantic analysis (pLSA) [21]. Each document is assumed to be a bag of words and is represented as a mixture of topics within the pLSA. Meanwhile, each topic is also a mixture of words in the vocabulary. Blei et al. (2003) assumed that both mixtures are a Dirichlet distribution and developed a Bayesian probabilistic topic model, i.e., the latent Dirichlet allocation (LDA) [22]. The two probabilistic topic models and their analogues have been extensively applied to analyze a variety of data, such as images [23], videos [24] and music [25].

When probabilistic topic models are employed to solve vision problems, visual features of natural images are commonly quantized as visual words using K-means, and each individual image is considered to be a document [26]. Generally speaking, the application of probabilistic topic models for vision problems consists of three common steps, as shown in Fig. 1(a): (1) generation of visual words, (2) model inference and (3) allocation of topic labels. Each natural image is modeled as a mixture of topics, which can be viewed as elements of a scene [27]. For example, an image of a sunset scene might be represented as a mixture of some topics, e.g., sky, mountain, water and so on. However, the topic models originally assumed that the data are exchangeable, i.e., the spatial relationship between visual words is ignored. Various approaches have been proposed to enhance the spatial consistency of topic labels. Sivic et al. (2005) used 'doublets' to encode spatially co-occurring regions [28, 29]. Wang et al. (2007) proposed the encoding of spatial information using document design tricks [26]. Verbeek et al. (2007) combined a MRF with topic models to determine the local regularization of label allocation [30, 31]. Larlus et al. (2010) combined a bag-of-words recognition with spatial regularization based on a mixture of both an MRF and the Dirichlet process [27].

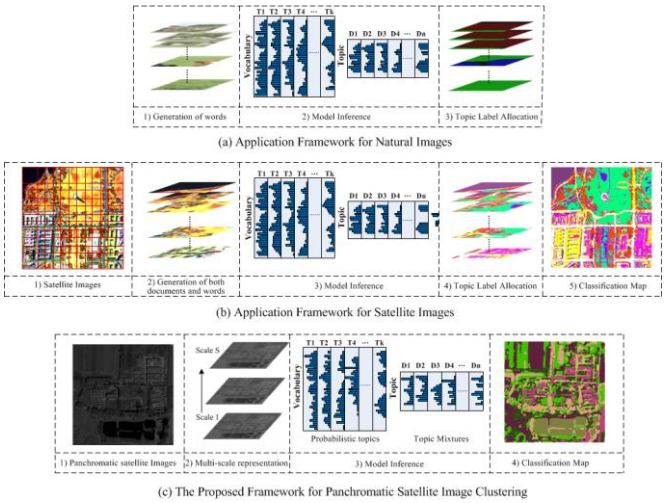
Probabilistic topic models have also emerged as a possible solution for discovering semantic patterns from VHR satellite images because they can simultaneously model both the distributions of geo-object classes (i.e., topics) and the co-occurrence between geo-objects in continuous image regions (i.e., topic distributions within documents). Ackay et al. (2008) combined the pLSA with a multi-scale segmentation algorithm for detecting geo-objects from VHR remote sensing images using an unsupervised framework [19]. Yi et al. (2011) presented a novel semantic clustering algorithm for VHR satellite images using the pLSA [20]. The LDA is also used to model each geo-category in a supervised framework, and then for annotating satellite images [32]. As shown in Fig. 1 (b), unlike the application of topic models to natural images, one needs to design both documents and visual words before using the topic models because there are too many geo-objects of interest in a large satellite image [16, 19-21, 32]. The satellite image is often partitioned into a set of sub-images (e.g., segments [19] or squared image areas [20, 32]) that are regarded as documents. In addition, to enforce the spatial consistency of the topic labels, the documents are commonly generated with some extent of overlap [16, 19-21, 32]. Consequently, the same pixel in the original satellite image could be allocated to multiple topic labels when it lies in different documents. Therefore, an additional step is often required to combine multiple topic labels for the same pixel to obtain a final classification map.

III. Multi-scale Latent Dirichlet Allocation

In this section, we first discuss major problems of geo-object clustering in VHR satellite images. Then, a multi-scale latent Dirichlet allocation (msLDA) is presented to solve the problems. Finally, an algorithm of the model is outlined.

A. Problem

Given a satellite image $\mathbf{I}$, let $\mathcal{T} = \{t = (i, j) \mid 1 \le i \le H, 1 \le j \le W)\}$ be the set of lattice sites in the image, where *W* and *H* are width and height of the image, respectively [33]. A random field indexed by the lattice system $\mathcal{T}$ is given by $\mathbf{X} = \{X_t = x_t \mid x_t \in \mathcal{A}, t \in \mathcal{T}\}$, where a random variable X_t at site *t* takes a value x_t in its state space $\mathcal{A} = \{1, 2, \ldots, d\}$. The set $\mathbf{x} = \{x_t \mid t \in \mathcal{T}\}$ is a sample drawn from the state space $\mathcal{X}$ with the joint



(a) Application Framework for Natural Images

(b) Application Framework for Satellite Images

(c) The Proposed Framework for Panchromatic Satellite Image Clustering

Fig. 1. Application framework of topic models for (a) Natural images, (b) Satellite images and (c) VHR panchromatic satellite images.

probability $P(\mathbf{X}=\mathbf{x})$. Given an observed sample $\mathbf{x}$ of a pixel attribute (e.g., spectrum or texture) random field $\mathbf{X}$, the goal of image clustering is to obtain a sample $\mathbf{z}=\{z_t \mid t\in\mathcal{T}\}$ of a label random field $\mathbf{Z}$ according to certain criterion, e.g. maximum a posterior $P(\mathbf{Z}\mid\mathbf{X})$.

Along with the increase of spatial resolution, some problems in satellite image clustering become increasingly serious: (1) the same geo-object with different spectra, (2) different geo-objects with nearly identical spectra and (3) salt-and-pepper effects in the clustering results. As shown in Fig. 1 (c), the two components in the proposed framework are aimed to solve the above-mentioned problems, i.e., the probabilistic topic model (including both probabilistic topics and topic mixture) and the multi-scale image representation.

In the proposed model, the statistical attribute of a geo-object is characterized by a probabilistic topic that is a multinomial mixture of words in the vocabulary. As shown in [22], the mixture could be used to describe the multimodal co-occurrence structures of words. Therefore, a geo-object with different spectra could be well characterized by a multimodal topic. Meanwhile, each document is represented as a mixture of topics. In other words, the co-occurrence relationship of geo-objects in a document can be characterized by the mixture of topics. Consequently, when different geo-objects with almost the same spectra frequently occur in a different scene, the mixture of topics could be employed to reflect the difference between them. In addition, a multi-scale image representation in the proposed framework provides a multi-scale viewpoint of each pixel because the grayscale value of a pixel in each image scale is used as a visual word, and a squared sub-image with a preset height h is regarded as a document covering the pixel. In the automatic framework, the clustering process is integrated into the model inference by both the document and scale selections to achieve the goal that multiple words on a lattice site are always allocated a unique topic label irrespective of the document in which they reside. Consequently, the salt-and-pepper effects are significantly reduced, and a self-adaptive spatial regularization effect for various geo-object categories could be achieved by maintaining a multi-scale spatial consistency.

B. Model

As shown in Fig. 1 (c), a multi-scale image representation of a panchromatic satellite image is derived by convoluting a given image with a variable-scale Gaussian [34]

$$L(x,y,s\delta)=G(x,y,s\delta)*I(x,y) \tag{1}$$

where * is the convolution operation in x and y, and the Gaussian $G(x,y,s\delta)=\frac{1}{2\pi(s\delta)^2}e^{-(x^2+y^2)/2s\delta}$ with scale s. Unlike Fig. 1 (b), it is not necessary to explicitly generate a collection of documents from a given satellite image because both the document and scale selections in the proposed framework are employed to ensure that multiple words on a lattice site are always allocated the same topic label irrespective of the document in which they reside. Therefore, we only need to implicitly know that there is a document with a preset size on each site. As shown in Fig. 2 (a), the msLDA assumes the following generative process for each site on the lattice system $\mathcal{T}$:

1) *Topic Sampling*

For each scale s, K topics are sampled from a Dirichlet prior $p(\vec{\phi}\mid\vec{\beta}_s)$, where each topic characterizes the grayscale histogram of a geo-object class and is a multinomial mixture of words in the vocabulary.

2) *Topic Mixture Sampling*:

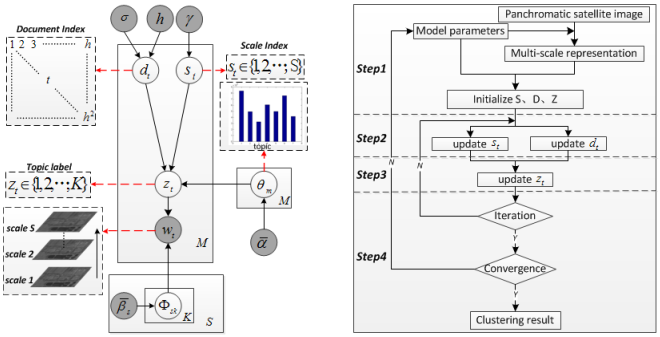
In the lattice system $\mathcal{T}$, a topic mixture vector for each site is sampled from a Dirichlet prior $p(\vec{\theta}\mid\vec{\alpha})$, where the topic mixture reflects the co-occurrence relationship among the geo-objects in a given document. Note that the topic mixture is independent of the scale in the multi-scale image representation because multiple words on a site share the same topic label.

3) *Scale Index Sampling*

In the multi-scale image representation, a scale index s_t is sampled for site t from a prior $p(s\mid\gamma)$. The scale index states from which scale the probabilistic topics are used to allocate a topic label to site t in the lattice system $\mathcal{T}$.

4) Document Index Sampling

In the lattice system $\mathcal{T}$, a document index d_t is sampled for site t from a normal distribution $p(d_t\mid\sigma)$, whose mean and variance are 0 and σ, respectively. The document index indicates the topic mixture vector would be employed for allocating a topic label to the site. Similar to the topic mixture vector, the document index is also independent of the image-representation scale.



(a) model (b) Algorithm

Fig. 2. The proposed model and algorithm. Some intuitional explanations of the model are shown at the end of red and dash arrows.

5) *Topic Label Sampling*

In the lattice system $\mathcal{T}$, topic label z_t of site t is sampled from a multinomial distribution $p(z \mid \vec{\theta})$, where the parameter vector $\vec{\theta}$ is the topic mixture vector of document d_t.

6) *Word Sampling*

In the lattice system $\mathcal{T}$, the grayscale value w_t of site t on s-th scale image is sampled from the sampled topic z_t with probability $p(w_t \mid z_t)$.

The key inferential problem in the msLDA is to compute the joint distribution of the observation and hidden variables on the lattice $\mathcal{T}$:

$$\begin{aligned} &P(\mathbf{W},\mathbf{Z},\mathbf{D},\mathbf{S} \mid \Phi,\Theta,\vec{\alpha},\vec{\beta},\gamma,\sigma,h) \\ &= P(\mathbf{W} \mid \mathbf{Z},\mathbf{D},\mathbf{S},\Phi,\vec{\beta})P(\mathbf{Z} \mid \mathbf{D},\mathbf{S},\Theta,\vec{\alpha})P(\mathbf{D} \mid \sigma,h)P(\mathbf{S} \mid \gamma) \end{aligned} \tag{2}$$

where $\mathbf{W}$ is the multi-scale representation of the source image $\mathbf{I}$; latent variable fields $\mathbf{Z}$, $\mathbf{D}$ and $\mathbf{S}$ are matrixes of topic labels, document indices and scale indices, respectively. Unlike the LDA, we assume that any site on the lattice $\mathcal{T}$ should be allocated the same topic label by the msLDA regardless of the document or scale to which it belongs. Therefore, the joint probability defined in Eq. (2) could be empirically approximated by a series of updating z_t, d_t and s_t

$$\begin{aligned} P(\mathbf{W},\mathbf{Z},\mathbf{D},\mathbf{S}) \propto \prod_{t\in\mathcal{T}} \{&P(w_t \mid \mathbf{W}_{t\neg},\mathbf{Z},\mathbf{D},\mathbf{S})P(z_t \mid \mathbf{Z}_{t\neg},\mathbf{W},\mathbf{D},\mathbf{S}) \\ &P(d_t \mid \mathbf{D}_{t\neg},\mathbf{W},\mathbf{Z},\mathbf{S})P(s_t \mid \mathbf{S}_{t\neg},\mathbf{W},\mathbf{Z},\mathbf{D})\} \end{aligned} \tag{3}$$

C. Algorithm

As shown in Fig. 2 (b), a Gibbs sampling algorithm is used to estimate the model parameters and approximate the posterior distribution of latent variables in the model. The algorithm includes four steps, as follows:

1) Step #1: Initializing or re-estimating the model parameters.

As shown in Fig. 2 (a), there are two types of parameters: (1) five scalars: number of scales S, number of topics K, height of sub-image (i.e., document) h, sigma σ of zero-mean normal distribution $p(d \mid \sigma)$ and mean of the uniformed prior $p(s \mid \gamma)$; (2) $S+1$ vectors: a K-dimensional Dirichlet parameter $\vec{\alpha}$ of $p(\vec{\theta} \mid \vec{\alpha})$ and S V-dimensional Dirichlet parameters $\vec{\beta}_s$ of $p(\vec{\phi} \mid \vec{\beta}_s)$ where V is the size of the vocabulary. All of scalar parameters are initialized once. The $S+1$ vector parameters would be re-estimated based on the estimated mixtures of topics in documents $p(\vec{\theta} \mid \vec{\alpha})$ and topics $p(\vec{\phi} \mid \vec{\beta}_s)$.

In addition, three matrices with the same size as the source panchromatic satellite image $\mathbf{I}$ are randomly initialized before they are updated as follows, i.e., matrix of the scale index $\mathbf{S}$, matrix of the document index $\mathbf{D}$, and matrix of the topic label $\mathbf{Z}$.

2) Step #2: Sampling both the scale and document indices for each site.

For each site t, we sample a new scale index s_t by approximating its posterior

$$P(s_t \mid \mathbf{S}_{t\neg},\mathbf{D},\mathbf{Z},\mathbf{W}) \propto P(s_t \mid \gamma)\frac{n^{(z_t)}_{t\neg,w^s_t} + \beta_{w^s_t}}{\sum_{w=1}^{N}(n^{(z_t)}_{t\neg,w^s} + \beta_{w^s})} \tag{4}$$

where $n^{(z_t)}_{t\neg,w^s_t}$ is the number of word w associated with topic z_t on the s-th scale, with the exception of the topic index on site t; β_{w^s} is w-th element of the Dirichlet parameter vector $\vec{\beta}_s$.

Meanwhile, for site t, we sample a new document index d_t by approximating its posterior

$$\begin{aligned} &P(d_t \mid \mathbf{D}_{t\neg},\mathbf{Z},\mathbf{S},\mathbf{W}) \\ &\propto \exp[\frac{(x_t - x^{d_t})^2 + (y_t - y^{d_t})^2}{\sigma}] * \frac{n^{(d_t)}_{t\neg,k} + \alpha_k}{\sum_{k'=1}^{K}(n^{(d_t)}_{t\neg,k'} + \alpha_{k'})} \end{aligned} \tag{5}$$

where (x_t, y_t) and (x^{d_t}, y^{d_t}) are the coordinates of site t and the center site of document d_t, respectively; $n^{(d_t)}_{t\neg,k}$ is the number of the k-th topic within document d_t, with the exception of the topic index on site t; α_k is the k-th element of the Dirichlet parameter vector $\vec{\alpha}$.

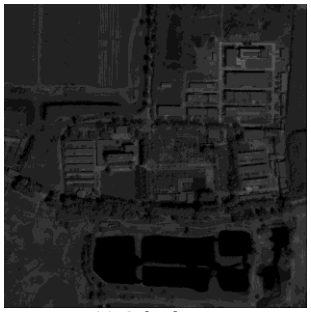
For the detailed derivation of Eq. (4) and (5), please refer to the Appendix.

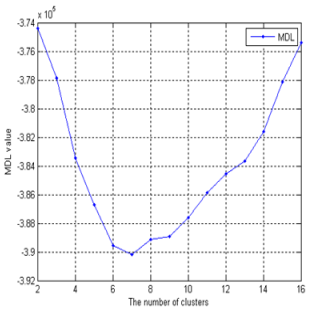
3) Step #3: Sampling a topic label for each site.

Given both the scale index, s_t and the document index d_t for site t, the posterior of topic label z_t is same as in the LDA

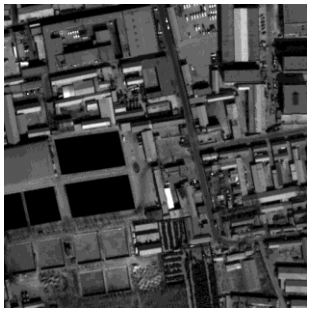
$$\begin{aligned} &P(z_t \mid \mathbf{Z}_{t\neg},\mathbf{D},\mathbf{S},\mathbf{W}) \\ &\propto \frac{n^{(d_t)}_{t\neg,k} + \alpha_k}{\sum_{k'=1}^{K}(n^{(d_t)}_{t\neg,k'} + \alpha_{k'})} * \frac{n^{(z_t)}_{t\neg,w^s_t} + \beta_{w^s_t}}{\sum_{w=1}^{N}(n^{(z_t)}_{t\neg,w} + \beta_{w^s})} \end{aligned} \tag{6}$$

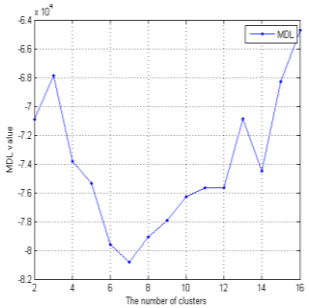
4) Step #4: Checking the convergence and driving the clustering results.


(a) Suburban



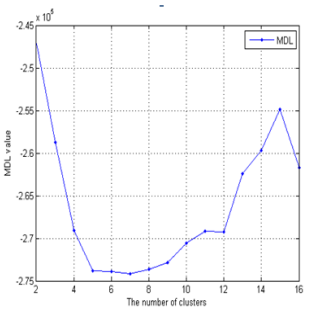
(b) Curve of value : suburban image


(c) Urban



(d) Curve of value for a urban image


(e) Rural



(f) Curve of value for a rural image

Fig. 3. Experimental datasets. (a), (c) and (e) are QUICKBIRD image of a suburb area, a QUICKBIRD image of an urban area and an EROS-B image of a rural area, respectively. (b), (d) and (f) are corresponding plots of MDL criterion.

After the Gibbs sampling procedure is completed, the algorithm determines whether convergence has been achieved. If convergence has not been achieved, the algorithm returns to step 1 to re-estimate the hyper-parameters of the model, i.e., $\vec{\alpha}$ and $\vec{\beta}_s$. Otherwise, the clustering result is derived by maximizing the posterior distribution of the topic label for each pixel using the estimated parameters.

IV. Experimental Results and Discussions

In this section, we first describe the experimental images and the quantitative evaluation methods for the experimental results. Then, we analyze the effect of parameters in the msLDA on the clustering results. Finally, we compare the performance of the msLDA with that of state-of-the-art methods for three typical of geographic scenes in terms of both qualitative and quantitative aspects.

The proposed msLDA algorithm is coded and implemented in a MATLAB environment. The parameter configurations are determined based on the different experimental conditions.

A. Experimental Images and Evaluation Methods

1) *Experimental Images*

As shown in Fig. 3 (a), (c) and (e), three 0.6-m resolution panchromatic satellite images are used in our experiments, i.e., a QUICKBIRD image of a suburban area with a size of 900×900 pixels, a QUICKBIRD image of an urban area with a size of 500×500 pixels, and an EROS-B image of a rural area with a size of 800×800 pixels. In the suburban image, there are six types of major geo-objects: *building*, *road*, *shadow*, *water*, *field and tree*. Five geo-object classes (i.e., *building*, *road*, *shadow*, *water* and *ground*) are scattered in the urban image. However, in the rural image, two geo-object classes (i.e., *field* and *tree*) cover most of the image. In addition, there are some *water*, *shadow* and *road* in the rural image. Note that two types of geo-objects with almost the same grayscale values occur in all three images, i.e., *water* and *shadow*.

2) *Experimental Evaluation Methods*

In our experiments, two quantitative criteria are employed in addition to visual inspection to evaluate the clustering results, i.e., entropy and the Kappa coefficient [35]. The former is an information theoretical criterion that measures the uncertainty of clusters given a ground truth class and the uncertainty of classes given a cluster. However, the entropy is independent of whether a cluster correctly corresponds to a class. Therefore, the Kappa coefficient is also used to measure the consistency between clusters and classes.

a) *Entropy*

Let h_{ck} be the number of pixels assigned to the cluster k within a ground truth class c, $h_c = \sum_{k=1}^{K} h_{ck}$ be the total number of pixels within a ground truth class c, and $h_k = \sum_{c=1}^{C} h_{ck}$ is denoted by the number of pixels assigned to cluster k, where K is the number of clusters and C is the number of ground truth classes.

The quality of a cluster is measured in terms of the homogeneity of the ground truth classes within the cluster, i.e., the cluster entropy. For the k-th cluster, the cluster entropy, E_k, is given by

$$E_k = -\sum_{c=1}^{C} \frac{h_{ck}}{h_k} \log \frac{h_{ck}}{h_k} \quad [35] \tag{7}$$

Therefore, the overall cluster entropy, $E_{cluster}$, is given by a weighted sum of the individual cluster entropies as

$$E_{cluster} = \frac{1}{\sum_{k=1}^{K} h_k} \sum_{k=1}^{K} h_k E_k \quad [35] \tag{8}$$

The cluster entropy reflects the quality of the individual clusters in terms of the homogeneity of the pixels in a cluster. However, the cluster entropy continues to decrease as the number of clusters increases. To address this problem, another entropy measure could be defined that measures how the pixels of the same class are represented by the various clusters

created.

Given a ground truth class c, the quality of a clustering result is measured in terms of the homogeneity of the cluster labels within the class, i.e., the class entropy

$$E_c = -\sum_{k=1}^{K} \frac{h_{ck}}{h_c} \log \frac{h_{ck}}{h_c} \text{ [35]} \quad (9)$$

Then, the overall class entropy E_{class}, is given by a weighted sum of individual class entropies as

$$E_{class} = \frac{1}{\sum_{c=1}^{C} h_c} \sum_{c=1}^{C} h_c E_c \text{ [35]} \quad (10)$$

The overall entropy could be defined as a linear combination of the class entropy E_{class}, and the cluster entropy $E_{cluster}$

$$E = \beta E_{class} + (1-\beta) E_{cluster} \text{ [35]} \quad (11)$$

where $\beta \in [0,1]$ is a weight for balancing the two measures [19] and β is set to 0.5 in the experiments. Generally speaking, a smaller overall entropy value indicates higher homogeneity.

b) *Kappa Coefficient*

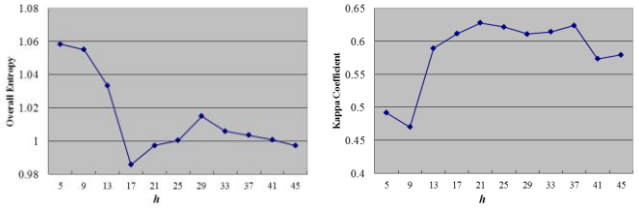
The Kappa coefficient (or Kappa statistic), which serves as a quantitative measurement of the agreement between the classification results and the ground truth maps, is one of the most commonly used statistics for evaluating the image classification accuracy. A Kappa of 1 indicates perfect agreement, whereas a Kappa of 0 indicates agreement that is equivalent to chance. In other words, the higher the value of Kappa, the higher the classification accuracy. To compute the Kappa coefficient, we use the minimum total number of misclassifications criterion to map the cluster labels to the ground truth labels [36].

B. *Effect of Parameters in the msLDA on Experimental Results*

As shown in Fig. 2 (a), there are two types of parameters in the msLDA: (1) four scalars: number of topics K, number of scales S, height of sub-images (i.e., documents) h, and the sigma σ of zero-mean normal distribution for document selection; (2) $S+1$ vectors: a K-dimensional Dirichlet parameter $\vec{\alpha}$ of $p(\vec{\theta} \mid \vec{\alpha})$ and S V-dimensional Dirichlet parameters $\vec{\beta}_s$ of $p(\vec{\phi} \mid \vec{\beta}_s)$ where V is the size of the vocabulary.

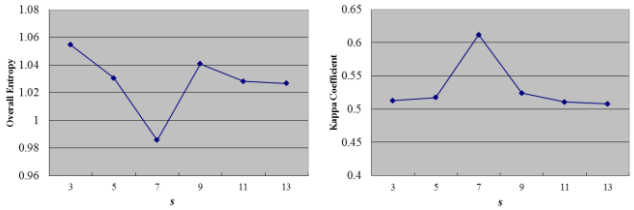
Among these parameters, the Dirichlet priors are initialized as symmetric priors and re-estimated during the model inference, i.e., $\alpha = 50/K$ and $\beta = 100$ [37]. The number of topics K, is estimated using a minimum description length (MDL) criterion [20, 32, 38]. As shown in Fig. 3 (b), (d) and (f), the optimal K corresponds to the minimal values in the MDL curves. For the detail of calculating the MDL value, please refer to [20, 32, 38]. In our experiments, the number of topics is set to 7.

In this subsection, we analyze the influence of three parameters: height h, of squared sub-images (i.e., size of documents), number of scales S, and the sigma of zero-mean normal distribution for document selection using the suburban image. Two special cases of the msLDA are also discussed.



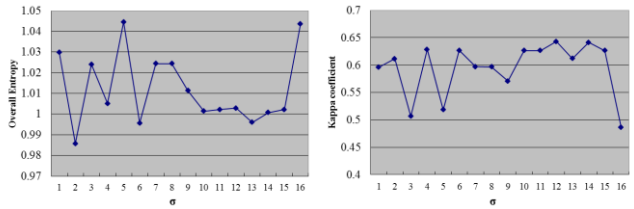
(a) (b)

Fig. 4. The quantitative performance evaluation of the clustering results. (a) Entropy versus sizes of documents. (b) The Kappa coefficient versus sizes of documents. The x-axes show the sizes of documents with h, which refers to the height of a square document.



(a) (b)

Fig. 5. The quantitative performance evaluation of the clustering results. (a) Entropy versus the number of scales. (b) The Kappa coefficient versus the number of scales.



(a) (b)

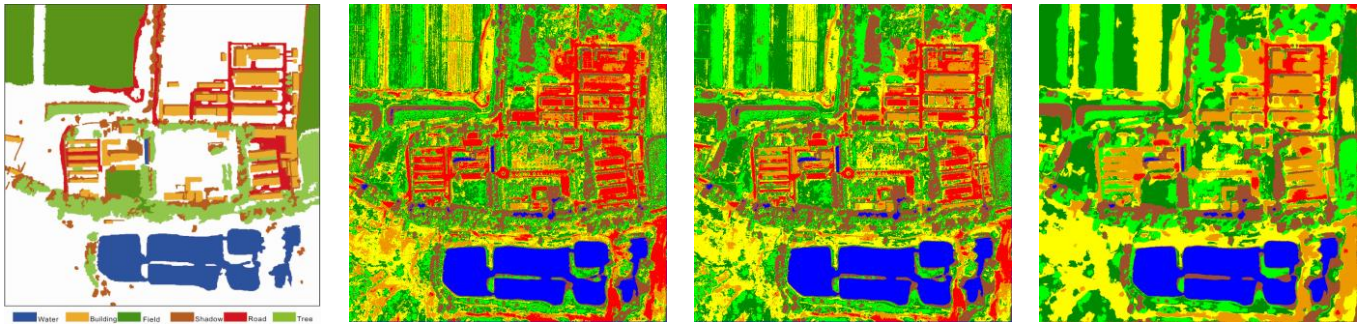
Fig. 6. The quantitative performance evaluation of the clustering results. (a) Entropy versus the value of sigma. (b) The Kappa coefficient versus the value of sigma.

1) *Size of Documents*

Given a site t in an image lattice, the size of documents determines two quantities: (1) the pixel that would be treated as the same document centered on site t; (2) the document that might be selected to sample the word (i.e., grayscale value) on site t. Fig. 4 shows the change in performance of the msLDA with varying sizes of documents. Both the entropy and the Kappa coefficient indicate that the performance of the msLDA increases with the size of documents at the beginning, then the performance becomes relatively stable when the size is larger than 17. Therefore, in the following experiments, the height of a sub-image is set to 17.

2) *Number of Scales*

The number of scales states how many times the original satellite images are convoluted. As shown in Fig. 5, the entropy is the lowest one, and the Kappa coefficient is the largest one when the number of the scales is equal to 7. In the following experiments, the number of scales is set to 7.



(a) Ground-truth (b) case#1: sigma =0, s=1 (c) case#2: sigma=2, s=1 (d) The msLDA clustering result

Fig. 7. Qualitative comparison of the multi-scale LDA clustering results with its two special cases. (a) Ground truth. (b) Case #1. (c) Case #2. (d) The msLDA.

3) *Sigma of Document Selection*

Given a site t in an image lattice, the sigma σ , of the normal distribution $p(d_t \mid \sigma)$ states the prior probabilistic distribution of documents, which is the probability that they might be used for sampling a latent topic label for site t . As shown in Fig. 6, there is not a significant variation in the trend of the model performance with the increase of the sigma σ . In the following experiments, we let σ equal to 2.

TABLE I
THE EVALUATION OF THE OVERALL ENTROPIES AND KAPPA COEFFICIENT OF DIFFERENT MODELS

Clustering models	The overall entropy	The Kappa coefficient
Case #1: sigma=0, s=1	1.07335	0.524058
Case #2: sigma=2, s=1	1.04484	0.569025
msLDA: sigma=2, s=7	0.98567	0.628443

4) *Special Cases of the msLDA*

In this subsection, we analyze two special cases of the msLDA for qualitatively and quantitatively revealing the characteristics of three complementary components by setting the model parameters: (1) case #1 topic model component: sigma = 0 and s = 1; (2) case #2 topic model component coupled with document selection: sigma = 2 and s = 1. In case #1, there is only one grayscale value on an image site t because s = 1, and its topic label is always sampled from the same document because of sigma = 0. Therefore, only the topic model component in the msLDA still works. In contrast, in case #2, although there is still one scale, all of the documents covering site t might be selected to sample a topic label for the word on site t .

For comparison, the clustering results of the two special cases coupled with that of the msLDA are shown in Fig. 7 (b), (c) and (d), respectively. As shown in Fig. 7 (b) and (c), there are two points worth noting: (1) the geo-object *road* is incorrectly clustered as the same label with almost all the instances of *building* in case #1; (2) part of *building* is successfully discriminated from *road* in case #2. Meanwhile, by comparing Fig. 7 (b), (c) and (d), we find that (1) almost all instances of *building* are correctly discriminated from *road* by the msLDA; and (2) the spatial consistency of cluster labels in the msLDA is significantly better than that of the two spatial cases.

As shown in Table I, the quantitative evaluations also indicate that the performance improves with the addition of the model components from case #1 (topic model) to case #2 (topic model + the novel framework with document selection), and the msLDA (topic model + the novel framework + multi-scale image representation).

C. Performance Evaluation over Satellite Images of Multiple Geographic Scenes

In this subsection, we compare the performance of the msLDA with that of K-means, ISOdata, the LDA (the special case #1 mentioned in the subsection IV-B; for simplicity, it is termed as the LDA) and the LDA+MRF [31] for three typical geographic scenes. We first analyze the spatial consistency of the clustering results for the suburban image when the topic labels are assumed to be an MRF. Then, the performance of the msLDA is compared with other clustering algorithms for both the urban and rural images.

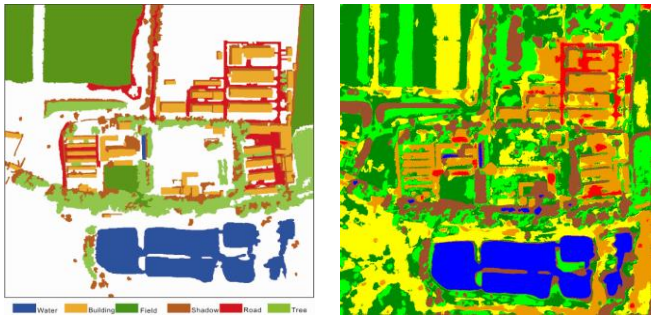
1) *Suburban Image*

Verbeek and Triggs (2007) introduced an MRF for topic labels to enhance the spatial consistency [31]. As shown in Fig. 1, the proposed application framework is significantly different from that where the topic model is employed to analyze natural images. To reveal the different characteristics of the msLDA from existing work, we borrowed the idea of [31] to assume an MRF over topic labels, which is termed the MRF+LDA. The algorithm includes two steps: (1) the topic model is learned as the LDA and (2) the clustering result is obtained by maximizing the energy function

$$E = E_d + \lambda E_s , \qquad (12)$$

where λ is a weight for balancing the a data energy E_d and a smoothness energy E_s [39].

The ground truth of suburban image and clustering results are shown in Fig. 8, where a series of clustering results are obtained for the LDA+MRF by changing the values of the smoothing parameter λ . On one hand, a small value of λ (e.g. λ = 1) does not exert a significant smoothing effect on the clustering result (as shown in Fig. 8 (f)). On the other hand, a large value of λ (e.g. λ = 10) yields an over-smoothed result.



(a) Ground-truth (b) The msLDA

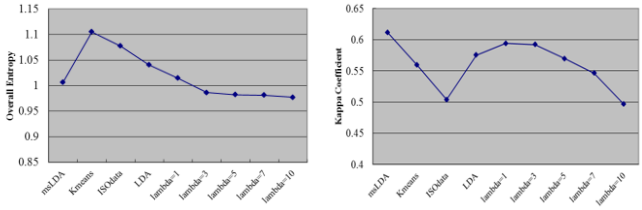
(c) K-means (d) ISOdata

(e) The LDA (f) The LDA+MRF with λ =1

(g) The LDA+MRF with λ =3 (h) The LDA+MRF with λ =5

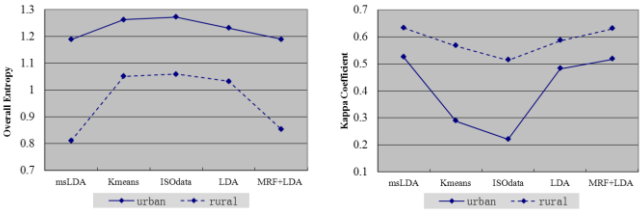
(i) The LDA+MRF with λ =7 (j) The LDA+MRF with λ =10

Fig. 8. Clustering results for a QUICKBIRD image of a suburban area. (a) Ground truth. (b) The msLDA. (c) K-means. (d) ISOdata. (e) The LDA. (f) - (j) are clustering results of the LDA+MRF with λ =1, λ =3, λ =5, λ =7 and λ =10, respectively.



(a) (b)

Fig. 9. Quantitative comparison of the clustering results for QUICKBIRD image of a suburban area. (a) Overall entropy for different methods. (b) The Kappa coefficient for different methods.

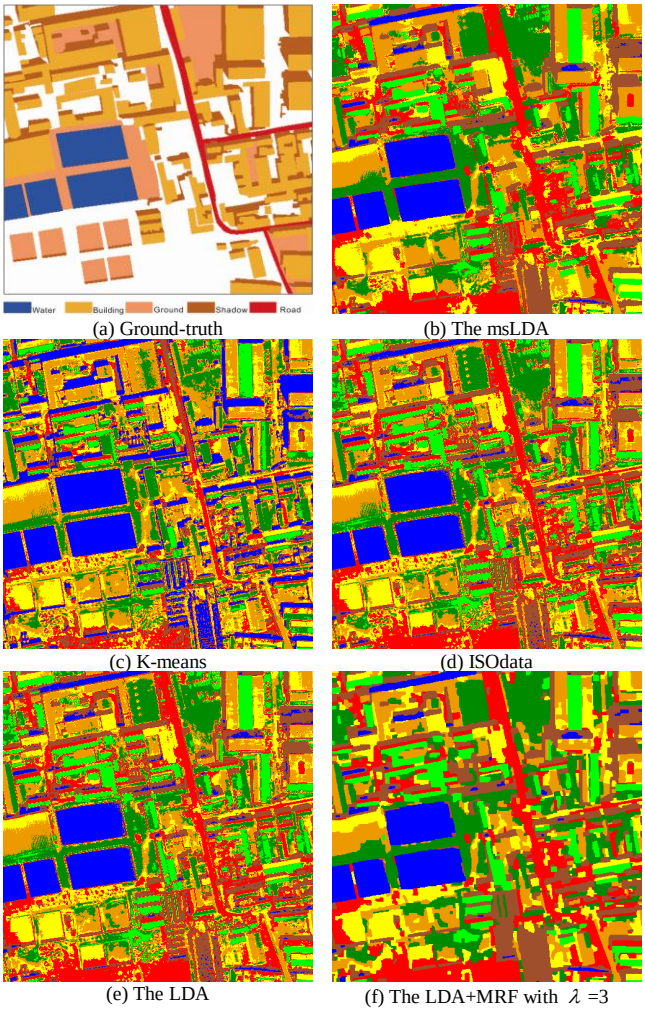


(a) (b)

Fig. 10. Quantitative comparison of the clustering results for a QUICKBIRD image of urban area and an EROS-B image of a rural area. (a) Overall entropy for different methods. (b) The Kappa coefficient for different methods

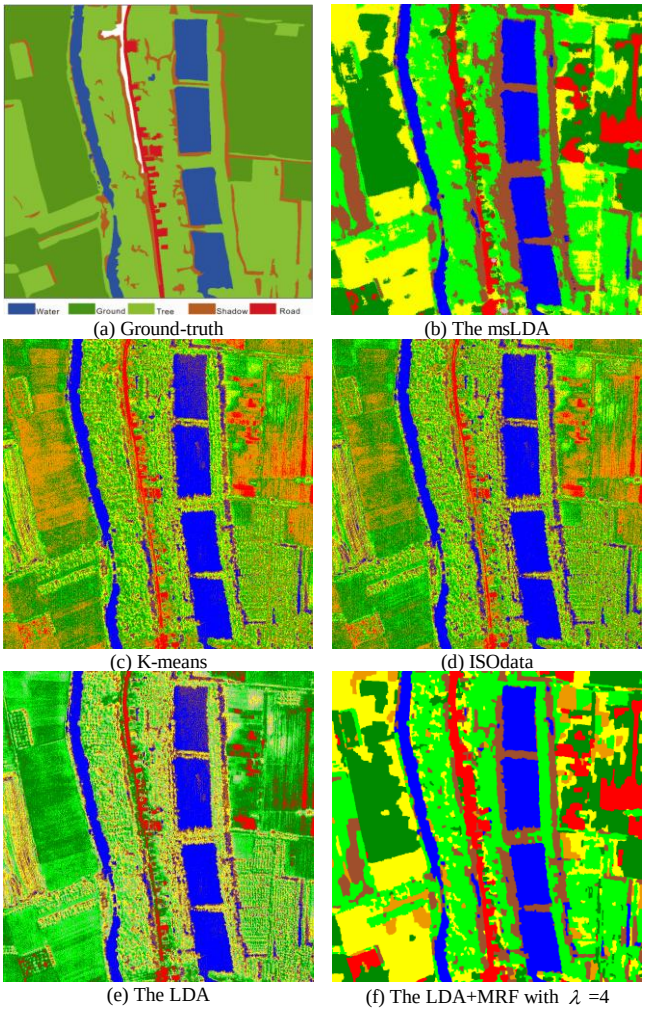
As illustrated in Fig. 8 (j), the large-scale *field* area located at the upper left corner of the image is smoothed appropriately. However, the modality of *road* in fine scale located in approximately the upper right corner is blurred with the neighboring *building*. The underlying reason is that the spatial smoothing of cluster labels is realized by the individual parameter λ, which imposes an equal smoothing penalty towards all the geo-objects in an image. Unlike the MRF scheme, through a multi-scale image representation coupled with scale selection, the msLDA could realize a self-adaptive smoothing effect on clustering results according to different geo-object types at different scales. As shown in Fig. 8, the field in a homogeneous area is heavily smoothed, and the edge of road remains well preserved due to a small amount of smoothing. In addition, irrespective of the value of λ, the error of incorrectly clustering the geo-object *road* into the same label as part of *building* cannot be corrected by assuming an MRF over labels. However, as shown in Fig. 8 (b), the mistake in the msLDA can be corrected by combining the topic model with a multi-scale image representation in the framework embedded with both document and scale selections.

From visual inspection, the spatial consistency of cluster labels in both the msLDA and the LDA+MRF clustering results (especially when $\lambda > 3$) is obviously better than that of both K-means and ISOdata. Furthermore, it can be seen from Fig. 8 (c) and (d) that in the clustering results of the K-means and ISOdata, almost all instances of *shadow* are incorrectly clustered into the same class as *water* because the grayscale values of both *shadow* and *water* are almost the same. In contrast, they are almost correctly separated into two different



(a) Ground-truth (b) The msLDA (c) K-means (d) ISOdata (e) The LDA (f) The LDA+MRF with λ =3

Fig. 11. Clustering results for a QUICKBIRD image of a urban area. (a) Ground truth. (b) The msLDA. (c) K-means. (d) ISOdata. (e) The LDA. (f) The LDA+MRF with λ =3.



(a) Ground-truth (b) The msLDA (c) K-means (d) ISOdata (e) The LDA (f) The LDA+MRF with λ =4

Fig. 12. Clustering results for an EROS-B image of a rural area. (a) Ground truth. (b) The msLDA. (c) K-means. (d) ISOdata. (e) The LDA. (f) The LDA+MRF with λ =4.

clusters by the topic-related models (e.g., LDA, LDA+MRF, msLDA). It is known that the clustering process of both K-means and ISOdata are actually conducted through the grayscale segmentation because every grayscale scale is treated independently. However, a set of neighboring pixels are regarded as a document and are modeled in the topic-related models. In other words, both grayscale values of the individual pixel and the neighboring spatial information have been used in the models.

For quantitative evaluation, on one hand, the msLDA is better than the other methods, as shown in Fig. 9 (a), in terms of the Kappa coefficient. On the other hand, the msLDA is not as good as that in the LDA+MRF (when λ =3, 5, 7, 10) in terms of the overall entropy, although it is still better than both the K-means and ISOdata. As shown in Fig. 9 (b), the larger the smoothing parameter, the lower the entropy in the MRF+LDA. Even when λ =10, the clustering result is obviously over-smoothed and not a desirable result. Nevertheless, the entropy in this clustering result becomes lower. Therefore, it might be necessary to resort to another performance measure. This is one of reasons for using the Kappa coefficient to reflect a desirable clustering result. Generally speaking, the msLDA compares favorably with other methods, particularly in the aspects of classification accuracy and adaptive smoothing effect.

2) *Both Urban and Rural Images*

It is important to emphasize that in this experiment only the best result of the LDA+MRF with optimal smoothing parameter λ is selected for comparison. As for the urban and rural images, the value of λ is equal to 3 and 4, respectively.

As shown in Fig. 10, the msLDA outperforms other clustering methods in terms of both overall entropy and Kappa coefficient. In addition, as shown in Figs. 11 and 12, although there are different geographic objects in the urban and rural scenes, the msLDA still distinguishes *shadow* from *water* very well; and the msLDA is significantly better than the K-means and ISOdata in terms of the spatial consistency of the cluster labels, and it is also better than the MRF+LDA.

V. Conclusions

In this paper, a multi-scale latent Dirichlet allocation model is proposed to address the problem of semantic clustering of geo-objects in VHR satellite panchromatic images. Our major contribution is to propose an automatic framework that combines a probabilistic topic model with a multi-scale image representation of a panchromatic satellite image. However, it is also possible to apply the proposed model to multi-spectral satellite images. For example, one might treat each band in a multi-spectral satellite image as a panchromatic image and add another constraint in which multiple spectral values on the same site always correspond to the same topic label. In the future, we will extend the proposed model to analyze multi-spectral images in a supervised learning framework.

Appendix Inference Using Gibbs Sampling

In this appendix, we derive the inference procedure in Gibbs sampling.

A. Updating d_t by Approximating $P(d_t \mid \mathbf{D}_{t\neg}, \mathbf{Z}, \mathbf{S}, \mathbf{W})$

Given $\mathbf{D}_{t\neg}$, $\mathbf{Z}$ and $\mathbf{S}$, d_t could be sampled by

$$\begin{aligned}
&P(d_t \mid \mathbf{D}_{t\neg}, \mathbf{Z}, \mathbf{S}, \mathbf{W}) \\
&= \frac{P(\mathbf{D}, \mathbf{Z}, \mathbf{S}, \mathbf{W} \mid \vec{\alpha}, \vec{\beta}, \gamma, \sigma, h)}{\sum_{i=1}^{h\times h} P(d_t = i, \mathbf{D}_{t\neg}, \mathbf{Z}, \mathbf{S}, \mathbf{W} \mid \vec{\alpha}, \vec{\beta}, \gamma, \sigma, h)} \\
&= \frac{P(\mathbf{W} \mid \mathbf{Z}, \mathbf{S}, \vec{\beta}) P(\mathbf{Z} \mid \mathbf{D}, \vec{\alpha}) P(\mathbf{D} \mid \sigma, h) P(\mathbf{S} \mid \gamma)}{P(\mathbf{W} \mid \mathbf{Z}, \mathbf{S}, \vec{\beta}) P(\mathbf{S} \mid \gamma) \sum_{i=1}^{h\times h} P(\mathbf{Z} \mid d_t = i, \mathbf{D}_{t\neg}, \vec{\alpha}) P(d_t = i, \mathbf{D}_{t\neg} \mid \sigma, h)} \\
&= \frac{P(\mathbf{Z} \mid \mathbf{D}, \vec{\alpha}) P(\mathbf{D} \mid \sigma, h)}{\sum_{i=1}^{h\times h} P(\mathbf{Z} \mid d_t = i, \mathbf{D}_{t\neg}, \vec{\alpha}) P(d_t = i, \mathbf{D}_{t\neg} \mid \sigma, h)} \\
&= \frac{\prod_{t' \in \mathcal{T}} P(z_{t'} \mid d_t, \vec{\alpha}) P(d_t \mid \sigma, h)}{(\sum_{i=1}^{h\times h} P(z_t \mid d_t = i, \vec{\alpha}) P(d_t = i \mid \sigma, h)) \prod_{t' \in \mathcal{T}_{t\neg}} P(z_t \mid d_t, \vec{\alpha}) P(d_t \mid \sigma, h)} \\
&= \frac{P(z_t \mid d_t, \vec{\alpha}) P(d_t \mid \sigma, h)}{\sum_{i=1}^{h\times h} P(z_t \mid d_t = i, \vec{\alpha}) P(d_t = i \mid \sigma, h)}
\end{aligned} \tag{13}$$

where $P(z_t \mid d_t, \vec{\alpha}) \propto \frac{P(\vec{z}_{d_t} \mid, \vec{d}_t, \vec{\alpha})}{P(\vec{z}_{d_{t\neg}} \mid \vec{d}_{d_{t\neg}}, s_t, \vec{\alpha})}$. Following the derivation in [40], $\frac{P(\vec{z}_{d_t} \mid, \vec{d}_t, \vec{\alpha})}{P(\vec{z}_{d_{t\neg}} \mid \vec{d}_{d_{t\neg}}, s_t, \vec{\alpha})} \propto \frac{n_{t\neg,k}^{(j)} + \alpha_k}{\sum_{k'=1}^{K} (n_{t\neg,k'}^{(j)} + \alpha_{k'})}$. In this paper, we assume the prior of document is a Gaussian distribution. Therefore, $P(d_t \mid \mathbf{D}_{t\neg}, \mathbf{Z}, \mathbf{S}, \mathbf{W})$ can be approximated by

$$\begin{aligned}
&P(d_t \mid \mathbf{D}_{t\neg}, \mathbf{Z}, \mathbf{S}, \mathbf{W}) \\
&\propto \exp\left(\frac{(x_t - x^{d_t})^2 + (y_t - y^{d_t})^2}{\sigma}\right) * \frac{n_{t\neg,k}^{(j)} + \alpha_k}{\sum_{k'=1}^{K} (n_{t\neg,k'}^{(j)} + \alpha_{k'})}
\end{aligned} \tag{14}$$

where (x_t, y_t) and (x^{d_t}, y^{d_t}) are coordinates of t-th pixel and the center pixel of the d_t sub-image, respectively.

B. Updating s_t by Approximating $P(s_t \mid \mathbf{S}_{t\neg}, \mathbf{D}, \mathbf{Z}, \mathbf{W})$

Given $\mathbf{S}_{t\neg}$, $\mathbf{D}$ and $\mathbf{Z}$, s_t can be sampled by

$$\begin{aligned}
&P(s_t \mid \mathbf{S}_{t\neg}, \mathbf{D}, \mathbf{Z}, \mathbf{W}) \\
&= \frac{P(\mathbf{W} \mid \mathbf{Z}, \mathbf{S}, \vec{\beta}) P(\mathbf{Z} \mid \mathbf{D}, \vec{\alpha}) P(\mathbf{D} \mid \sigma, h) P(\mathbf{S} \mid \gamma)}{P(\mathbf{Z} \mid \mathbf{D}, \vec{\alpha}) P(\mathbf{D} \mid \sigma, h) \sum_{s=1}^{S} P(\mathbf{W} \mid \mathbf{Z}, s_t = s, \mathbf{S}_{t\neg}, \vec{\beta}) P(s_t = s, \mathbf{S}_{t\neg} \mid \gamma)} \\
&= \frac{P(w_t \mid z_t, s_t, \vec{\beta}) P(s_t \mid \gamma)}{\sum_{s=1}^{S} P(w_t \mid z_t, s_t = s, \vec{\beta}) P(s_t = s \mid \gamma)} \\
&= \frac{\frac{P(\vec{w}_{d_t} \mid \vec{z}_{d_t}, s_t, \vec{\beta})}{P(\vec{w}_{d_{t\neg}} \mid \vec{z}_{d_{t\neg}}, s_t, \vec{\beta})} P(s_t \mid \gamma)}{\sum_{s=1}^{S} \frac{P(\vec{w}_{d_t} \mid \vec{z}_{d_t}, s_t = s, \vec{\beta})}{P(\vec{w}_{d_{t\neg}} \mid \vec{z}_{d_{t\neg}}, s_t = s, \vec{\beta})} P(s_t = s \mid \gamma)} \\
&\propto P(s_t \mid \gamma) \frac{n_{t\neg,w_t^*}^{(z_t)} + \beta_{w_t^*}}{\sum_{w=1}^{N} (n_{t\neg,w}^{(z_t)} + \beta_w)}
\end{aligned} \tag{15}$$

Acknowledgments

This work is supported by the National Basic Research Program of China (No. 2011CB707102), National Natural Science of China (No. 40901217), the Program for New Century Excellent Talents in University (No. NECT-11-0039) and the Fundamental Research Funds for the Central Universities (No. 105563GK).

References

[1] A. Paoli, F. Melgani, and E. Pasolli, "Clustering of hyperspectral images based on multiobjective particle swarm optimization," IEEE Transactions on Geoscience and Remote Sensing, vol. 47, pp. 4175-4188, Dec. 2009.

[2] D. Lu and Q. Weng, "A survey of image classification methods and techniques for improving classification performance," International Journal of Remote Sensing, vol. 28, pp. 823-870, Mar. 2007.

[3] A. Mukhopadhyay and U. Maulik, "Unsupervised pixel classification in satellite imagery using multiobjective fuzzy clustering combined with SVM classifier," IEEE Transactions on Geoscience and Remote Sensing, vol. 47, pp. 1132-1138, Apr. 2009.

[4] A. P. Doulgeris, S. N. Anfinsen, and T. Eltoft, "Automated non-Gaussian clustering of polarimetric synthetic aperture radar images," IEEE Transactions on Geoscience and Remote Sensing, vol. 49, pp. 3665-3676, Oct. 2011.

[5] G. J. McLachlan and D. Peel, Finite mixture models. New York: John Wiley & Sons, 2000.

[6] Y. Zhang, M. Brady, and S. Smith, "Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm," IEEE Transactions on Medical Imaging, vol. 1, pp. 45-57, Jan. 2001.

[7] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 6, pp. 721-741, Nov. 1984.

[8] C. Nikou, A. Likas, and N. Galatsanos, "A Bayesian framework for image segmentation with spatially varying mixtures," IEEE Transactions on Image Processing, vol. 19, pp. 2278-2289, Sep. 2010.

[9] C. Nikou, N. P. Galatsanos, and A. C. Likas, "A class-adaptive spatially variant mixture model for image segmentation," IEEE Transactions on Image Processing, vol. 16, pp. 1121-1130, Apr. 2007.

[10] J. L. Marroquin, E. A. Santana, and S. Botello, "Hidden Markov measure field models for image segmentation," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 25, pp. 1380-1387, Nov. 2003.

[11] G. S. Xia, C. He, and H. Sun, "Integration of synthetic aperture radar image segmentation method using Markov random field on region adjacency graph," Radar, Sonar & Navigation, IET, vol. 1, pp. 348-353, Oct. 2007.

[12] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Spectral–spatial hyperspectral image segmentation using subspace multinomial logistic regression and Markov random fields," IEEE Transactions on Geoscience and Remote Sensing, vol. 50, pp. 809-823, Mar. 2012.
[13] S. Martinis, A. Twele, and S. Voigt, "Unsupervised extraction of flood-induced backscatter changes in SAR Data using Markov image modeling on irregular graphs," IEEE Transactions on Geoscience and Remote Sensing, vol. 49, pp. 251-263, Jan. 2011.
[14] F. Melgani and S. B. Serpico, "A Markov random field approach to spatio-temporal contextual image classification," IEEE Transactions on Geoscience and Remote Sensing, vol. 41, pp. 2478-2487, Nov. 2003.
[15] T. Blaschke, "Object based image analysis for remote sensing," ISPRS Journal of Photogrammetry and Remote Sensing vol. 65, pp. 2-16, Jan. 2010.
[16] D. Blei, L. Carin, and D. Dunson, "Probabilistic topic models," IEEE Signal Processing Magazine, vol. 27, pp. 55-65, Nov. 2010.
[17] G. Castilla and G. J. Hay, "Image objects and geographic objects," in Object-based image analysis: spatial concepts for knowledge-driven remote sensing applications, T. Blaschke, S. Lang, and G. J. Hay, Eds., ed: Springer, 2008.
[18] S. Deerwester, S. T. Dumais, G. W. Furnas, T. K. Landauer, and R. Harshman, "Indexing by latent semantic analysis," Journal of the American Society for Information Science, vol. 41, pp. 391-407, Sep. 1990.
[19] H. G. Akcay and S. Aksoy, "Automatic detection of geospatial objects using multiple hierarchical segmentations," IEEE Transactions on Geoscience and Remote Sensing, vol. 46, pp. 2097-2111, Jul. 2008.
[20] W. Yi, H. Tang, and Y. Chen, "An object-oriented semantic clustering algorithm for high resolution remote sensing images using the aspect model," IEEE Geoscience and Remote Sensing Letter, vol. 8, pp. 522-526, May 2011.
[21] T. Hofmann, "Probabilistic latent semantic indexing," in Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Berkeley, CA, USA, Aug. 15-19, 1999, pp. 50-57.
[22] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," Journal of Machine Learning Research, vol. 3, pp. 993-1022, Jan. 2003.
[23] D. M. Blei and M. I. Jordan, "Modeling annotated data," in Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Informaion Retrieval, Toronto, ON, Canada, Jul. 28-Aug. 1, 2003, pp. 127-134.
[24] Y. Wang and G. Mori, "Human action recognition by semi-latent topic models," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 31, pp. 1762-1774, Oct. 2009.
[25] M. D. Hoffman, D. M. Blei, and P. R. Cook, "Finding latent sources in recorded music with a shift-invariant HDP," in Proceedings of the 12th International Conference on Digital Audio Effects(DAFx-09), Como, Italy, Sep. 1-4, 2009.
[26] X. Wang and E. Grimson, "Spatial latent Dirichlet allocation," in Proceedings of Neural Information Processing Systems Conference (NIPS) 2007, Vancouver, B.C., Canada, Dec. 3-5, 2007, pp. 1577-1584.
[27] D. Larlus, J. Verbeek, and F. Jurie, "Category level object segmentation by combining bag-of-words models with Dirichlet processes and random fields," International Journal of Computer Vision, vol. 88, pp. 238-253, Jun. 2010.
[28] J. Sivic, B. C. Russell, A. A. Efros, A. Zisserman, and W. T. Freeman, "Discovering objects and their location in images," in Proceedings of the 2005 IEEE International Conference on Computer Vision, Beijing, China, Oct. 17-21, 2005, pp. 370-377.
[29] J. Sivic, B. C. Russell, A. Zisserman, W. T. Freeman, and A. A. Efros, "Unsupervised discovery of visual object class hierarchies," in Proceedings of the 2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Anchorage, Alaska, USA, Jun. 23-28, 2008, pp. 1-8.
[30] B. Zhao, F.-F. Li, and E. P. Xing, "Image segmentation with topic random field," in Proceedings of the 11th European conference on Computer vision: Part V, Crete, Greece, Sep. 5-11, 2010.
[31] J. Verbeek and B. Triggs, "Region classification with Markov field aspect models," in Proceedings of the 2007 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Minneapolis, MN Jun. 17-22, 2007.
[32] M. Li énou, H. Ma îre, and M. Datcu, "Semantic annotation of satellite images using latent Dirichlet allocation," IEEE Geoscience and Remote Sensing Letter, vol. 7, pp. 28-32, Jan. 2010.
[33] H. Deng and D. A. Clausi, "Unsupervised image segmentation using a simple MRF model with a new implementation scheme," Pattern Recognition, vol. 37, pp. 2323-2335,, Apr. 2004.
[34] T. Lindeberg, "Scale-space theory: a basic tool for analysing structures at different scales," Journal of Applied Statistics (Supplement on Advances in Applied Statistics: Statistics and Images: 2), vol. 21, pp. 224-270, 1994.
[35] M. Halkidi, Y. Batistakis, and M. Vazirgiannis, "On clustering validation techniques," Journal of Intelligent Information Systems, vol. 17, pp. 107-145, Dec. 2001
[36] J. Cohen, "A coefficient of agreement for nominal scales," Educational and Psychological Measuremnt, vol. 20, pp. 37-46, Apr. 1960.
[37] H. M. Wallach, D. Mimno, and A. McCallum, "Rethinking LDA: why priors matter," in Proceedings of Neural Information Processing Systems Conference (NIPS) 2009, Vancouver, B.C., Canada, Dec. 7-9, 2009, pp. 1973-1981.
[38] I. O. Kyrgyzov, O. O. Kyrgyzov, H. Ma îre, and M. Campedel, "Kernel MDL to determine the number of clusters," in Proceedings of the 5th international conference on Machine Learning and Data Mining in Pattern Recognition, Leipzig, Germany, Jul. 18-20, 2007, pp. 203-217.
[39] R. Szeliski, R. Zabih, D. Scharstein, O. Veksler, V. Kolmogorov, A. Agarwala, M. Tappen, and C. Rother, "A comparative study of energy minimization methods for Markov random fields with smoothness-based priors," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 30, pp. 1068-1080, Jun. 2008.
[40] G. Heinrich, "Parameter estimation for text analysis," Technical Note, Aug. 2008.

Hong Tang received his B.S. and M.S. degrees from the China University of Mining and Technology, Xuzhou, China, in 1998 and 2001, respectively, and his Ph.D. in Pattern Recognition and Intelligence Systems from the Shanghai Jiao Tong University, Shanghai, China, in 2006.

From Mar. 2006 to Mar. 2008, he worked as a post-doc at the IMEDIA project of INRIA Paris. He is currently an Associate Professor at the Beijing Normal University. His research interests include remote-sensing image processing, pattern recognition, and natural disaster reduction.

Li Shen received his B.S. degree in Photogrammetry and Remote Sensing from Wuhan University, Wuhan, China, in 2008.

He is currently pursuing his Ph.D. at the College of Resources Science and Technology, Beijing Normal University, China. His research focuses on remote sensing image processing and analysis.

Yinfeng Qi received the B.S. degree in Geographic Information System from Southeast University, Nanjing, China, in 2010.

She is now pursuing the M.S. degree at the Academy of Disaster Reduction and Emergency Management, Beijing Normal University, China. Her research mainly focuses on sensitivity analysis of parameters involved in probabilistic topic models.

Yunhao Chen received his B.S. and M.S. degrees in Resource Management from the Anhui University of Science and Technology, Huainan, China, in 1994 and 1997, respectively, and his Ph.D. in Geodesic Engineering from the China University of Mining and Technology, Beijing, in 1999.

He is currently a Professor at the Beijing Normal University, Beijing, and a Special Research Scholar of the National Disaster Reduction Committee, Ministry of Civil Affairs of China. From 2000 to 2001, he did postdoctoral research at the Beijing Normal University. From 2001 to the present, he has been with the College of Resource Science and Technology, Beijing Normal University. His research interests include thermal

remote sensing applications in urban heat island phenomena, evapotranspiration, and natural disaster reduction.

Yang Shu received his B.S. degree in Geographic Information Systems from the Southeast University, Nanjing, China, in 2009.

He is currently pursuing his Ph.D. at the Academy of Disaster Reduction and Emergency Management, Beijing Normal University, China. His research concentrates on remote sensing image processing and analysis.

Jing Li received his B.S. degree in Geography and his M.S. degree in Graphics and Remote Sensing from Peking University, Beijing, China, in 1982 and 1985, respectively.

He is currently the Director of the Institute of Resource Technology and Engineering, College of Resource Science and Technology, Beijing Normal University, Beijing. From 1990 to 1991, he was a Visiting Scholar at the University of Hawaii. From 1993 to 2003, he was with the Institute of Remote Sensing and Geographical Information Systems, Peking University. From 2003 to the present, he has been a Professor with the Beijing Normal University and Chief Engineer of the National Disaster Reduction Committee, Ministry of Civil Affairs of China. His research interests include remote sensing of natural resources, remote sensing applications in natural disaster reduction, and monitoring of oceanic and coastal ecology by remote sensing.

David A. Clausi earned his B.A.Sc. (1990), M.A.Sc. (1992), and Ph.D. (1996) in Systems Design Engineering at the University of Waterloo (Waterloo, Ontario, Canada).

After completing his doctorate, Prof. Clausi worked in software medical imaging at Agfa (Waterloo). He started his academic career in 1997 as an Assistant Professor in Geomatics Engineering at the University of Calgary, Canada. In 1999, he returned to his alma mater and is now a Professor specializing in the fields of Intelligent and Environmental Systems as well as the departmental Graduate Chair. Prof. Clausi is an active interdisciplinary and multidisciplinary researcher. He has an extensive publication record, publishing refereed journal and conference papers in the diverse fields of remote sensing, computer vision, algorithm design, and biomechanics. His research efforts have led to successful commercial implementations, including the building and selling of his own company. He was the Co-chair of IAPR Technical Committee 7 - Remote Sensing during 2004-2006.

He has received numerous scholarships, paper awards, and two Teaching Excellence Awards. In 2010, he received the award for "Research Excellence and Service to the Research Community" by the Canadian Image Processing and Pattern Recognition Society (CIPPRS).