SPARSE CODING-BASED TOPIC MODEL FOR REMOTE SENSING IMAGE SEGMENTATION

Jun Shi *Zhiguo Jiang* *Hao Feng* *Yibing Ma*

Image Processing Center, School of Astronautics
Beijing University of Aeronautics and Astronautics
100191, Beijing, China
Beijing Key Laboratory of Digital Media
100191, Beijing, China

ABSTRACT

Land cover segmentation can be viewed as topic assignment that the pixels are grouped into homogeneous regions according to different semantic topics in topic model. In this paper, we propose a novel topic model based on sparse coding for segmenting different kinds of land covers. Different from conventional topic models which generally assume each local feature descriptor is related to only one visual word of the codebook, our method utilizes sparse coding to characterize the potential correlation between the descriptor and multiple words. Therefore each descriptor can be represented by a small set of words. Furthermore, in this paper probabilistic Latent Semantic Analysis (pLSA) is applied to learn the latent relation among word, topic and document due to its simplicity and low computational cost. Experimental results on remote sensing image segmentation demonstrate the excellent superiority of our method over *k*-means clustering and conventional pLSA model.

Index Terms—remote sensing, sparse coding, pLSA, land cover segmentation

1. INTRODUCTION

Land cover segmentation plays a key role on remote sensing image understanding, which generally groups the image pixels or the low-level feature descriptors (color, texture, shape, etc) into semantically homogeneous regions. The segmented regions can provide much convenience for object detection, environment perception and other applications. However, the pixels or the low-level features fail to precisely describe high-level semantic concepts which may exist in the remote sensing image.

To narrow down the gap between the low-level features and high-level semantics, the topic model, *e.g.*, probabilistic latent semantic analysis (pLSA) [1] and latent Dirichlet allocation (LDA) [2], has been proposed and successfully applied to remote sensing images [3,4]. Topic model usually assumes that low-level features can be modeled as words and images are viewed as documents. It aims to discover the latent topics hidden between words and documents, which are used to represent the image. Specifically, topic model firstly performs Bag-of-Words (BoW) scheme [5] that each local feature descriptor is represented by one visual word of a codebook through vector quantization (VQ). Consequently, the distribution of words among documents can be gained. Then the latent topics will be learnt by the co-occurrence property of words through a probabilistic graphical model. With the assumption that each feature descriptor is related to only one visual word in BoW strategy, the potential correlation between the descriptor and multiple words is ignored. Therefore, the representation of words generated by BoW is likely to be not optimal and lack of discriminant information. Recently, sparse coding (SC) [6,7] has attracted much attention, which generally makes use of a relatively small set of words from the codebook to represent each feature descriptor. It not only minimizes the reconstruction error, but also preserves the sparsity of feature representation. More importantly, different from BoW strategy, SC considers the relationships that may exist between the descriptor and multiple words.

In this paper, we introduce a sparse coding based topic model in land cover segmentation, which can explore the potential correlation between each feature descriptor and multiple words and thus can generate a fine encoding of words. Concretely, the combination of Gabor feature and RGB color is firstly taken as the feature descriptor of each pixel, and then sparse coding is applied to obtain the word representation of each pixel. Note that each image is blocked into non-overlapping regions which are viewed as the documents of topic model. Furthermore, considering

This work was supported by the National Natural Science Foundation of China (No. 61071137, 61071138, 61027004) and the 973 Program of China (No. 2010CB327900).

pLSA is simpler and has lower computational cost than LDA, pLSA model is applied to learn the topics hidden between words and documents. Finally the words are grouped into different semantic topics learnt by pLSA. Experimental results on remote sensing image segmentation demonstrate our method outperforms original *k*-means clustering and conventional pLSA model.

The rest of the paper is arranged as follows: in Section 2 we give a brief overview of sparse coding and probabilistic latent semantic analysis. The proposed method is introduced in Section 3. Section 4 presents the experimental results and finally we conclude this paper in Section 5.

2. SPARSE CODING AND PROBABILISTIC LATENT SEMANTIC ANALYSIS

2.1. Sparse Coding (SC)

Assume that $\mathbf{X}$ is a set of M D-dimensional local feature descriptors from an image, *i.e.*, $\mathbf{X}=[\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_M]\in\mathbf{R}^{D\times M}$, and $\mathbf{Y}$ is the codebook including N visual words, *i.e.*, $\mathbf{Y}=[\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N]\in\mathbf{R}^{D\times N}$. Conventional Bag-of-Words (BoW) [5] strategy generally represents one local feature descriptor by one word. Thus the potential correlation between the descriptor xi and other words is likely to be ignored.

To address the problem, sparse coding (SC) [6,7] performs sparsity regularization term to enhance the ability of word representation. The SC problem can be written as Eq. (1), where $\mathbf{c}_i$ is the code of $\mathbf{x}_i$ and λ is the regularization parameter [4]. Note that the item $\|\mathbf{c}_i\|_1$ is the l_1-norm of $\mathbf{c}_i$, which is used to characterize the sparsity of the code $\mathbf{c}_i$.

$$\arg\min_{\mathbf{Y},\mathbf{c}_i}\left\{\sum_{i=1}^{M}\|\mathbf{x}_i - \mathbf{Y}\mathbf{c}_i\|^2+\lambda\|\mathbf{c}_i\|_1\right\}. \qquad (1)$$

2.2. Probabilistic Latent Semantic Analysis (pLSA)

As a generative model, pLSA [1] aims to cluster words into topics based on their co-occurrence property. Specifically, the graphical model representation of pLSA is shown in Fig. 1.

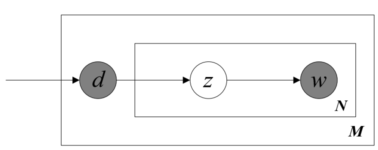


Fig. 1 Graphical model representation of pLSA

Given a collection of M documents $\mathbf{D}=[d_1, d_2, \ldots, d_M]$, and each document has N_i words. Nodes d and w are the observed document and word, and node z is the latent topic. Therefore, a joint probability model $P(w,d)$ can be written as follows:

$$P(w,d)=P(d)P(w\mid d)=P(d)\sum_z P(w\mid z)P(z\mid d)\,, \qquad (2)$$

The learning problem of pLSA can be expressed by a maximum likelihood formulation as shown in Eq. (3), where $n(d,w)$ denotes how often the word w occurred in the document d and K is the number of topics,

$$\begin{aligned} L &= \sum_{i=1}^{M}\sum_{j=1}^{N} n(d_i,w_j)\log P(d_i,w_j) \\ &= \sum_{i=1}^{M} n(d_i)\left[\log P(d_i)+\sum_{j=1}^{N}\frac{n(d_i,w_j)}{n(d_i)}\log\sum_{k=1}^{K}P(w_j\mid z_k)P(z_k\mid d_i)\right] \end{aligned} \qquad (3)$$

3. SPARSE CODING-BASED TOPIC MODEL

3.1. Feature Description

Considering remote sensing images usually contains plenty texture and structure information, we use the Gabor filter responses under different scales and orientations to characterize each pixel. Consequently, the variation of the texture and structure information can be detected. Moreover, RGB values of different kinds of land covers are also employed for describing the color information. Finally each pixel can be represented by the combination of Gabor features and RGB colors.

3.2. Sparse Coding-based Topic Model

On the basis of the low-level feature representation, *i.e.*, Gabor and RGB features, the topic model is applied to discover the latent high-level semantics in remote sensing images. In this paper, we make use of sparse coding (SC) to represent each descriptor instead of BoW strategy. SC representation can be regarded as continuous and also can discover some existed correlation between the descriptor and multiple words. Based on the continuous representation for words, pLSA is employed for grouping words into topics. Consequently, each word is assigned by one latent topic, and the regions with similar texture and color information can be obtained. The segmentation framework of our method is presented in Fig. 2 and the overall process is described as follows:

1) 1) Extract Gabor and RGB features for each pixel. Note that the Gabor filter responses with 4 scales and 6 orientations are used as texture feature descriptor. Thus each pixel can be represented by a 27 dimensional feature vector.

2) 2) Perform *k*-means clustering to construct the codebook. Block the whole image into non-overlapping regions, and the regions are similar to the documents of pLSA model. The SC representation of each pixel in the region is obtained through Eq. (1), and the feature-sign search algorithm [8] is the solution to the SC problem. Finally max pooling scheme [6,7] is utilized to pool the SC codes of each pixel from the region together and then sum normalization [6,7] is applied to normalize the pooled feature, which can be viewed as probability $P(w|d)$.

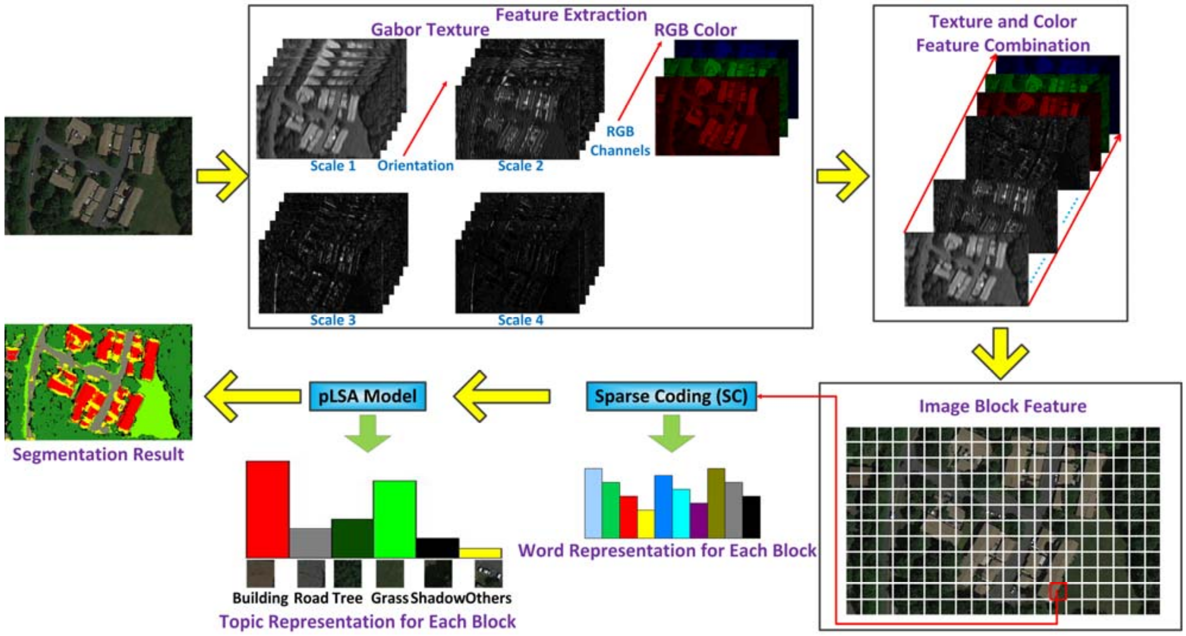


Fig. 2 Segmentation framework of our method

3) Calculate the probability $P(w|z)$ through the Expectation Maximization (EM) algorithm [1], and thus the words can be grouped into topics, and the whole image can also be segmented into different kinds of land covers according to homogeneous topics.

4. EXPERIMENTAL RESULTS

In this experiment, we use the high resolution remote sensing images with the size of 500×300 to evaluate the performance of our method, and these images are labeled into 7 semantic concepts, *i.e.*, road, ground, tree, grass, building, shadow and others (including vehicles, edges and other textures). Specifically, the size of non-overlapping regions is 10×10, and the number of words generated by *k*-means clustering is set to 150, in addition, 9 topics are employed for indicating the semantics of the words. The two methods, *i.e.*, *k*-means clustering and conventional pLSA based on Gabor and RGB features, are used for comparison. Note that the regions of images are manually labeled into different classes for ground truth. Table 1 demonstrates the accuracies of three methods.

As can be seen from Table 1, pLSA model basically outperforms *k*-means clustering since it considers the high-level semantic features. Compared with *k*-means clustering and pLSA model, our method achieves excellent performance for important land covers, *i.e.*, ground, grass, tree and building. It may be explained by that the continuous word representation generated by sparse coding contributes to learning the topics that can describe the characteristics of land covers well. The segmentation results for remote sensing images are shown in Fig. 3, and different colors indicate different semantic concepts. Fig. 3(a) displays some original remote sensing images, and the result of *k*-means clustering is exhibited in Fig. 3(b). The topic assignments of pLSA and our method are respectively shown in Fig. 3(c) and Fig. 3(d). Specifically, the blue and red areas denotes the building, and forest green areas means the trees, and grass green areas can be viewed as the grass. Road is expressed by the gray color. Clearly heterogeneous regions can be better segmented by our method.

Table 1. Accuracies for segmentation of three methods (%)

Classes	*k*-means clustering	pLSA	Our method
Road	64.77	**80.04**	78.87
Ground	25.24	63.66	**68.94**
Grass	80.53	61.44	**81.89**
Tree	50.22	57.62	**75.87**
Building	82.20	70.87	**86.82**
Shadow	**72.30**	71.48	68.32
Others	17.65	38.28	**49.57**

5. CONCLUSIONS

In this paper, a novel topic model based on sparse coding for remote sensing image segmentation is proposed, which combines sparse coding and pLSA model. It makes use of sparse coding to obtain the word representation of each feature descriptor and then constructs the pLSA model to learn the semantic topics. Consequently, it not only takes into account the potential correlation between each descriptor and multiple words, but also narrows the gap between the low-level features and high-level semantics. Experiments on remotes sensing image segmentation demonstrate the effectiveness and feasibility of our method. Moreover, in the future work, we will combine sparse coding with spatial relation of words to further improve the segmentation results.

6. REFERENCES

[1] T. Hofmann, "Unsupervised learning by probabilistic latent semantic analysis," *Mach. Learn*. 42(1-2), pp. 177-196, 2001.

[2] D.M. Blei, A.Y. Ng, and M.I. Jordan, "Latent dirichlet allocation," *Journal of Machine Learning Research*, vol. 3, pp. 993-1022, 2003.

[3] Wenbin Yi, Hong Tang and Yunhao Chen, "An Object-Oriented Semantic Clustering Algorithm for High-Resolution Remote Sensing Images Using the Aspect Model," *IEEE Geoscience and Remote Sensing Letters*, vol. 8, no. 3, pp. 522-526, May 2011.

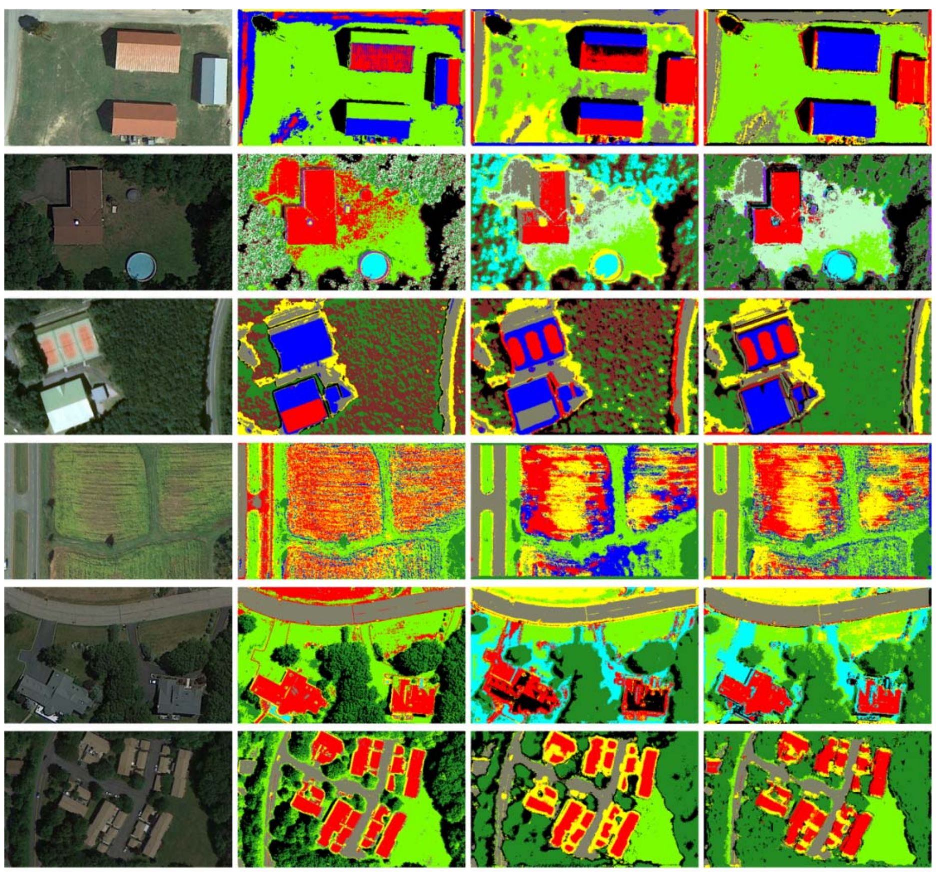
[4] Hao Feng, Zhiguo Jiang, and Xingmin Han, "Texture segmentation for remote sensing image based on texture-topic model," *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 2669-2672, 2011.

[5] Fei-Fei L. and Perona P., "A bayesian hierarchical model for learning natural scene categories," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* pp.524-531, 2005.

[6] Jianchao Yang, Kai Yu, Yihong Gong, and T. Huang, "Linear spatial pyramid matching using sparse coding for image classification," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1794-1801, 2009.

[7] J.J. Wang, J.C. Yang, K. Yu, F.J. Lv, T.S. Huang, and Y. H. Gong, "Locality-constrained linear coding for image classification," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* pp.3360-3367, 2010.

[8] Honglak Lee, Alexis Battle, Rajat Raina, and Andrew Y. Ng. "Efficient sparse coding algorithms," *Advances in Neural Information Processing Systems (NIPS)* 19, 2007.



(a) (b) (c) (d)

Fig.3 Segmentation results. (a) original image, (b) results of *k*-means clustering, (c) results of pLSA, (d) results of our method.