EARTH OBSERVATION IMAGE SEMANTICS: LATENT DIRICHLET ALLOCATION BASED INFORMATION DISCOVERY

Reza Mohammadi Asiyabi[1], Mihai Datcu[1,2]

[1]Research Center for Spatial Information (CEOSpaceTech), University POLITEHNICA of Bucharest (UPB), Bucharest, Romania
[2]Earth Observation Center (EOC), German Aerospace Center (DLR), Wessling, Germany

IGARSS 2021 - 2021 IEEE International Geoscience and Remote Sensing Symposium | 978-1-6654-0369-6/21/$31.00 ©2021 IEEE | DOI: 10.1109/IGARSS47720.2021.9553122

ABSTRACT

Land cover maps are among the most important products of Remote Sensing (RS) imagery. Despite remarkable advancements in land cover classification techniques, abundant detailed information in the very high-resolution RS images necessitates further improvements to harness the data and discover detailed semantic information. Moreover, scarcity of the labelled data and its quality is a major limitation in RS land cover mapping. In the present study, Latent Dirichlet Allocation is employed for semantic discovery in RS images and a novel kernel-based Bag of Visual Words model is proposed for land cover mapping.

Index Terms—Latent Dirichlet Allocation, Semantic Discovery, Bag of Visual Words, Topic Modeling, Classification

1. INTRODUCTION

Identification of land cover establishes baseline information for various activities. Land cover mapping from satellite imagery is among the most important applications of Remote Sensing (RS) and many researches have proposed different methodologies for this purpose [1]–[3]. Recent advancements in sensor technology have provided unprecedented amount of very high-resolution satellite images with abundant information about the land surface; therefore, it is necessary to improve semantic analysis methods in order to harness the latent semantic content of the images and produce accurate classification maps.

Generally, classification methods can be applied either unsupervised or supervised. The main problem with unsupervised classifiers is the lack of semantic meaning of the classes. On the other hand, supervised classifiers, despite providing semantically meaningful classes, require Ground Truth (GT) data for training. In addition to the high cost and unavailability of GT data for many case studies, the user-defined GT data is not sufficiently reliable. Besides, due to the complexity of the land cover in very high-resolution RS images, many semantic classes might get neglected and as a result, misclassified in other classes.

In order to resolve this issue, data mining techniques for semantic discovery of RS images are useful. Unsupervised semantic discovery methods such as Latent Dirichlet Allocation (LDA) can be applied to RS images to identify potential semantic classes. The user can interpret the identified potential semantic classes to produce or enhance the existing GT data and improve the classification results.

LDA is a generative probabilistic model in text analysis that has been proposed by Blei et al. [4] and has been utilized for latent semantic classes (topics) discovery in image processing. In [5], LDA was used to discover the latent structure of Synthetic Aperture Radar (SAR) images for application-oriented content classification in the explainable machine learning framework. LDA has been used in [6] for semantic annotation of large satellite images. A structural topic model, based on LDA, was developed in [2] for unsupervised latent topic feature extraction from high-resolution RS images with prior text knowledge.

Moreover, mid-level data representations (e.g., Bag of Visual Words (BOVW)) are capable of enhancing image classifiers [3], [7]. BOVW model is originated from Bag of Words model in text mining and has been utilized for image classification in many researches. Codewords distribution learning is used in [8] for natural scene categorization. Li et al. [3] supplemented the BOVW model by affinity propagation clustering algorithm and spatial pyramid matching technique to classify Polarimetric SAR data. And Bahmanyar et al. [1] applied LDA on BOVW to develop a more semantically interpretable model, Bag of Topics.

In the present study, LDA is used for semantic information discovery and a kernel-based BOVW model is proposed for pixel-wise representation of RS images. For this purpose, LDA is applied to the kernel-based BOVW histograms to generate the topic maps and used for discovering potential semantic classes. Finally, Support Vector Machine (SVM) classifier is used to classify the RS image with the proper semantic classes.

2. METHODOLOGY

Figure 1. illustrates the flowchart of the proposed methodology. In the first step, low-level features, including three spectral bands in the visible portion of the electromagnetic spectrum, are extracted from the RS image. Similar to the conventional BOVW model, visual dictionary is constructed using the well-known k-means clustering algorithm.

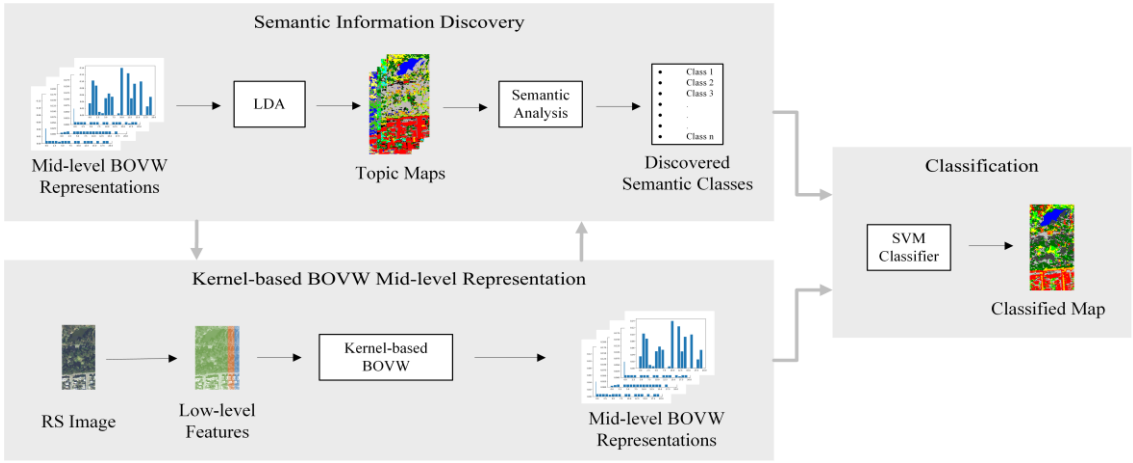

Figure 1. Flowchart of the proposed methodology. In the Kernel-based BOVW stage, the pixel-wise mid-level representation of the RS image is produced using the proposed kernel-based BOVW model. The resulting histograms are used in the Semantic Information Discovery stage to discover potential semantic classes using the LDA model. Finally, BOVW histograms and the discovered semantic classes are utilized to produce the pixel-wise classified map.

Through test and trial, visual dictionary size is set to 20 visual words, considering the kernel size and image specifications. A very high-resolution RGB image of the study area is used as the Ground Truth (GT) and a limited number of semantic classes are identified through visual interpretation. Training and testing samples are labelled according to the generated GT map. For the training stage, a 15×15 weighted kernel is centered on each training sample and the whole 225 pixels covered by the kernel are used to construct the BOVW histogram through coding and pooling steps. The weighted kernel is used to increase the effect of the closer pixels and the size of the kernel is set to 15×15 to obtain meaningful histograms. The constructed histogram is assigned to the center pixel and is considered as the BOVW representation of the pixel. Train BOVW histograms are utilized to train the SVM classifier with the RBF kernel. The trained SVM classifier is used to classify the image using the BOVW histograms of the whole scene.

Furthermore, BOVW histograms are fed into the LDA model to produce the topic maps with different numbers of topics. The resulting topic maps are used in parallel with the very high-resolution RGB image to analyze latent semantic information, discover neglected semantic classes and correct the GT map. New train and test samples from the new GT map are used in the kernel-based BOVW model to construct new BOVW histograms and finally producing the final classified map with the corrected semantic classes.

3. EXPERIMENTAL RESULTS AND DISCUSSION

A subset of very high-resolution RGB image from USGS aerial imagery [9] over San Francisco Bay, USA acquired in September 2008 was used in this study. The USGS RGB image has a spatial resolution of 0.3 m [9].

3.1. Semantic Discovery

In the first step, five semantic classes including buildings, roads, vegetation, water, and shadow are visually identified and the GT map is created. Kernel-based BOVW histograms are constructed according to the explained methodology in section 2 and the classified map is produced. The RGB image with the user-defined GT for 5 semantic classes and the resulting classified map are illustrated in Figure 2.

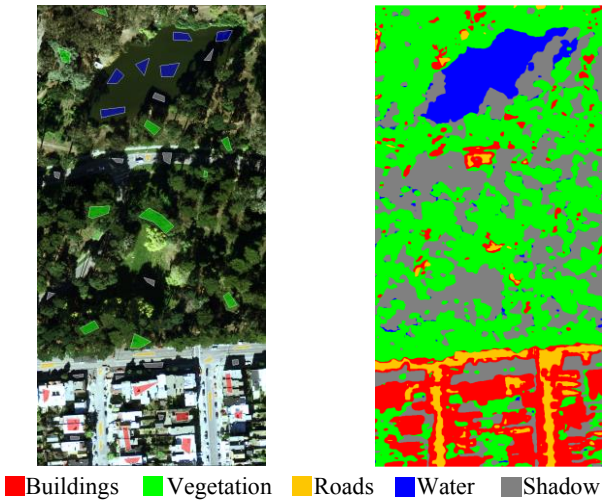

Figure 2. USGS RGB aerial image with labeled GT data and classified map with 5 semantic classes.

In the next step, BOVW histograms are fed into the LDA topic model and resulting topic maps with 5, 8, 10, and 20 topics are produced. LDA is a probabilistic topic model and the output is a vector of T probability values, where T denotes the number of the topics. The highest probability has been chosen as the topic label of the pixel to produce the topic maps. Resulting topic maps are illustrated in Figure 3.

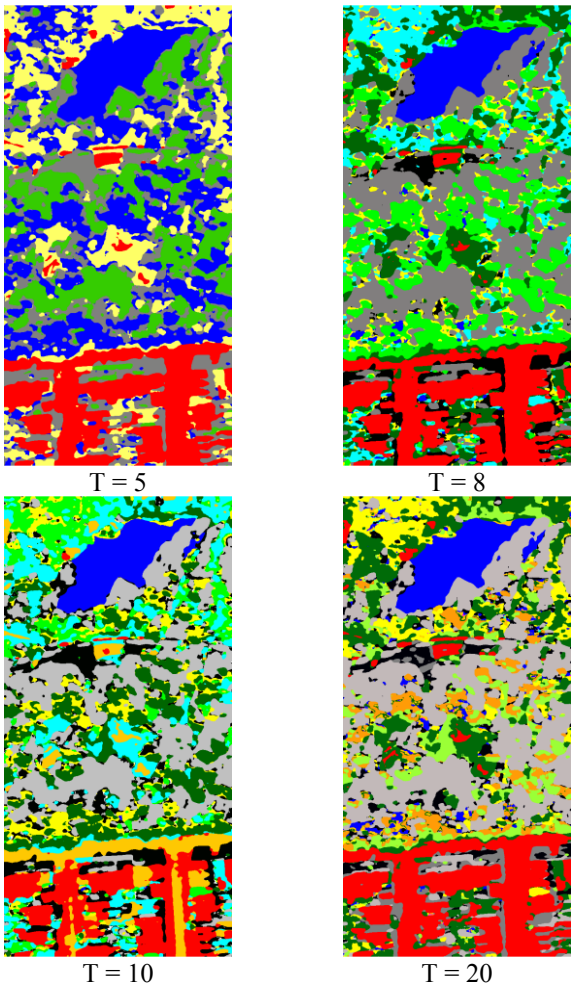

Figure 3.LDA topic maps with various topic numbers

Comparing the classified map and topic map with 5 topics (same as the number of the classes), some differences between the user-defined semantic classes and topics identified by the LDA model can be noticed. First of all, LDA tends to combine the buildings and roads into a single topic. Besides, water regions and very dark green vegetation are represented by the same topic. Another topic seems to be the representation of the sparse vegetation, bare soil and some trees with light green color. However, LDA tends to separate shadows with a dark background such as dense vegetation and water, and shadows with a lighter background such as buildings or roads.

When the number of topics in the LDA model is increased to 8, water region is represented by a separate topic and it is less mixed with vegetation areas. In addition, some other sparse vegetation covers are also got separated from other vegetation classes which results in four different topics representing vegetation covers, two shadow topics, one water topic and one topic for the constructed area.

With 10 topics, LDA represents buildings and roads by different topics. Water region and Shadows are still represented by one and two topics, respectively. The main difference is in vegetation covered areas where five different topics are identified by LDA. Even though different vegetation types including dense dark green vegetation, sparse and light green vegetation, sparse grass, and bare soil seem to be distinguished, but many small segments of vegetation are present in the topic map which seem to be redundant. It should also be noted that LDA with 10 topics was able to spot some vegetation under the shadows.

Increasing the number of topics to 20 does not seem to provide more information than 10 topics. But the redundancy of small segments with different topics representing different vegetation types has increased. It is observable on the top left or central parts of the topic map with 20 topics, illustrated in Figure 3.

3.2. GT Correction and Classification

Considering the findings of the previous analyses, new training and testing samples are labeled in 8 semantic classes including Buildings (B), Roads (R), Green trees (G), Yellowish vegetation (Y), Sparse vegetation (S), Water (W), Dark shadow (D), and Light shadow (L); and the classification result is generated through the proposed kernel-based BOVW model. The transformation between the initial user-defined 5 semantic classes and the corrected 8 semantic classes, and the RGB image with the GT regions for 8 semantic classes and the resulting classified map are illustrated in Figures 4 and 5, respectively.

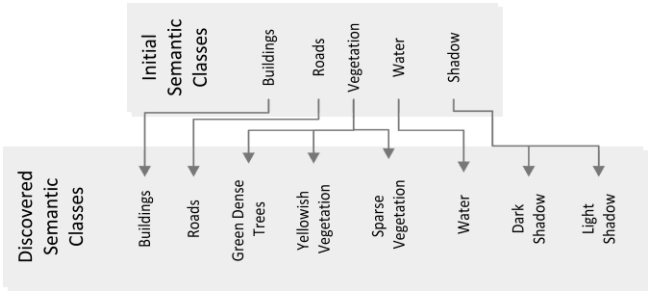

Figure 4. Transformation between the semantic classes.

Comparing the classified maps with 5 and 8 semantic classes, it is noticeable that 8-class map better distinguished buildings and roads. However, some sparse vegetation covered pixels are wrongly classified as buildings in the 8-class map, especially on the top left corner of the scene. This misclassification is also shown in the confusion matrix in Table 1. In addition, shadows with dark and light backgrounds are well distinguished in 8-class map, even though some yellowish vegetations are also classified as the light shadow. Moreover, dark green trees are well classified but sparse vegetation areas are underemphasized by being labelled as buildings and dark green trees in some areas. The confusion matrix demonstrates that the most confusing class is spars vegetation which is mixed with other vegetation and also constructed regions. Besides, roads and buildings got mixed due to the similarity of their color and construction material, as well as two semantic classes for shadow.

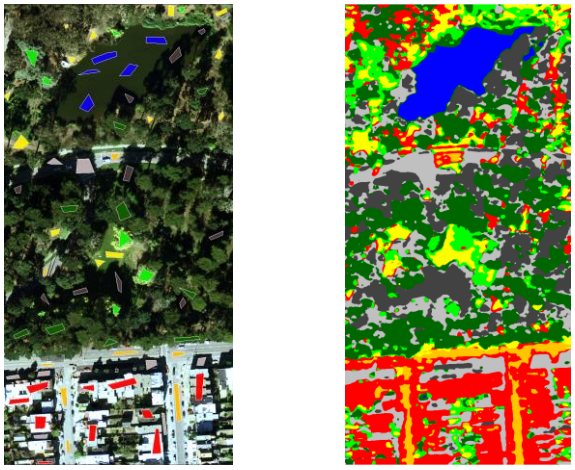

Figure 5. USGS RGB aerial image with corrected GT labels and classified map with 8 semantic classes.

Table 1. Normalized confusion matrix of the 8-class map.

	B	R	G	Y	S	W	D	L
B	94.3%	5.4%	0.0%	0.0%	0.2%	0.0%	0.0%	0.1%
R	11.4%	84.9%	0.0%	0.0%	3.7%	0.0%	0.0%	0.0%
G	0.0%	0.0%	99.6%	0.0%	0.3%	0.0%	0.1%	0.0%
Y	0.1%	0.3%	0.1%	81.2%	18.3%	0.0%	0.0%	0.0%
S	1.1%	1.6%	0.0%	10.2%	87.1%	0.0%	0.0%	0.0%
W	0.0%	0.0%	0.0%	0.0%	0.0%	100.0%	0.0%	0.0%
D	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	97.7%	2.3%
L	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	2.9%	97.1%

4. CONCLUSION

The detailed and complex information in the Very high-resolution RS images not only necessitates advanced representation and classification methods, but also demands an accurate and semantically comprehensive labelled GT map. The LDA topic model has been applied in this study to adjust the GT map based on semantic analysis of the generated topic maps with different topic numbers. In addition, a kernel-based BOVW model has been proposed for the pixel-wise representation of very high-resolution RS images. Contributions and conclusions of this study can be summarized as follows:

- Due to the complexity of the data in very high-resolution RS images, the user-defined semantic class labels are not comprehensive. As a result, many semantic classes might get neglected in the classified maps based on the user-defined labelled data.
- Topic models such as LDA are capable of discovering latent semantic information in very high-resolution RS images and can be utilized to adjust the labelled GT data and enhance the classified maps.
- The proposed kernel-based BOVW model achieved satisfactory classification results. However, further investigation is necessary in future studies to evaluate the performance of the kernel-based BOVW model as a pixel-wise alteration for patch-based BOVW model.

5. ACKNOWLEDGEMENTS

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 860370. Valuable scientific comments of Prof. Daniela Coltuc that immensely improved the manuscript is greatly appreciated. The authors would like to thank USGS [9] for making the data, used in this study, publicly available.

6. REFERENCES

[1] R. Bahmanyar, S. Cui, and M. Datcu, "A comparative study of bag-of-words and bag-of-topics models of EO image patches," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 6, pp. 1357–1361, 2015.

[2] H. Shao, Y. Li, Y. Ding, Q. Zhuang, and Y. Chen, "Land Use Classification Using High-Resolution Remote Sensing Images Based on Structural Topic Model," *IEEE Access*, vol. 8, pp. 215943–215955, 2020.

[3] X. Li, L. Zhang, L. Wang, and X. Wan, "Effects of BOW model with affinity propagation and spatial pyramid matching on polarimetric SAR image classification," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 10, no. 7, pp. 3314–3322, 2017.

[4] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *J. Mach. Learn. Res.*, vol. 3, no. Jan, pp. 993–1022, 2003.

[5] C. Karmakar, C. O. Dumitru, G. Schwarz, and M. Datcu, "Feature-free Explainable Data Mining in SAR Images Using Latent Dirichlet Allocation," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, 2020.

[6] M. Lienou, H. Maitre, and M. Datcu, "Semantic annotation of satellite images using latent Dirichlet allocation," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 1, pp. 28–32, 2009.

[7] Q. Zhu, Y. Zhong, B. Zhao, G.-S. Xia, and L. Zhang, "Bag-of-visual-words scene classifier with local and global features for high spatial resolution remote sensing imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 6, pp. 747–751, 2016.

[8] L. Fei-Fei and P. Perona, "A bayesian hierarchical model for learning natural scene categories," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 2005, vol. 2, pp. 524–531.

[9] U.S. Geological Survey, "USGS High Resolution Orthiomagery (Entity ID:1289214_10SEG445790)." https://earthexplorer.usgs.gov/ (accessed Sep. 30, 2020).