THE LOW BACKSCATTERING OBJECTS CLASSIFICATION IN POLSAR IMAGE BASED ON BAG OF WORDS MODEL USING SUPPORT VECTOR MACHINE

L. Yang [1] [*], L. Shi [1], P. Li [1], J. Yang [1], L. Zhao [2], B. Zhao [1]

[1] State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, China –
yang.le@whu.edu.cn
[2] School of Remote Sensing and Information Engineering, Wuhan University, China

Commission ICWG III/IVb

KEY WORDS: Low Backscattering Objects, Bag of Words, SIFT features, Support Vector Machine

ABSTRACT:

Due to the forward scattering and block of radar signal, the water, bare soil, shadow, named low backscattering objects (LBOs), often present low backscattering intensity in polarimetric synthetic aperture radar (PolSAR) image. Because the LBOs rise similar backscattering intensity and polarimetric responses, the spectral-based classifiers are inefficient to deal with LBO classification, such as Wishart method. Although some polarimetric features had been exploited to relieve the confusion phenomenon, the backscattering features are still found unstable when the system noise floor varies in the range direction. This paper will introduce a simple but effective scene classification method based on Bag of Words (BoW) model using Support Vector Machine (SVM) to discriminate the LBOs, without relying on any polarimetric features. In the proposed approach, square windows are firstly opened around the LBOs adaptively to determine the scene images, and then the Scale-Invariant Feature Transform (SIFT) points are detected in training and test scenes. The several SIFT features detected are clustered using K-means to obtain certain cluster centers as the visual word lists and scene images are represented using word frequency. At last, the SVM is selected for training and predicting new scenes as some kind of LBOs. The proposed method is executed over two AIRSAR data sets at C band and L band, including water, bare soil and shadow scenes. The experimental results illustrate the effectiveness of the scene method in distinguishing LBOs.

1. INTRODUCTION

Polarimetric Synthetic Aperture Radar (PolSAR), with its all-weather, all-time advantage, has been widely used in land-use mapping, change monitoring, postearthquake collapse assessment and so on (Zhao, 2014; Shi, 2015; Liu, 2017; Zhao, 2017). Image classification, as the basis of image interpretation, is one of the important applications of PolSAR(Lee, 2009). The prevalent classification methods are usually based on the polarization scattering information extracted from PolSAR images (Lang, 2012). Using the scattering entropy H and the average scattering angle α obtained from the H/α decomposition, the ground objects can be divided into eight scattering mechanisms to realize the unsupervised classification of PolSAR image(Lee, 1999). The Wishart classifier is introduced to the H/α classification method soon afterwards, that is, combining both polarimetric and statistical characteristics, to overcome the problem of excessively arbitrary boundary and improve the classification accuracy by performing Wishart iteration on the results of the H/α method(Cloude, 1997).

The LBOs, with similar polarimetric scattering characteristics and low backscattering intensity, are always confused in common classification methods that are based on polarimetric features or depend on the image intensity (Lang, 2012; Shi, 2012; Zhao, 2013; Pulvirenti, 2014). Some efforts have been made in the LBOs discrimination. Shi. et al. proposed the phase-diff standard deviation (PSD) between HH and VV channel and distinguished water, soil and shadow in the entropy-PSD space plane(Shi, 2012). Lang. et al. introduced the new defined entropy and anisotropy based on Freeman three-component decomposition and chose proper threshold to extract water and shadow from the whole image (Lang, 2012). Zhao. et al. used the statistical method, combining H-α target decomposition and

modified likelihood ratio test, to classify road, soil and shadow (Zhao, 2013).

However, the polarimetric characteristics of LBOs is usually unstable, leading to failure of the methods mentioned above in other experiment data sets. To settle the problem, we don't think about polarimetric features any more. Ant it can be noticed that the surrounding scene details of the LBOs are very overt and abundant. For instance, the radar shadow is always near the mountain; and the water channel is usually narrow and appears in town. Hence, this paper will make full use of the scene information around the LBOs, without introducing any polarimetric characteristic, to complete the discrimination. The experiment results of AIRSAR full polarimetric data sets based on the SVM classifier prove the effectiveness of the proposed method.

2. METHODLOGY

Scene image classification, defined as observing what the given image contains and then determining its category, has a hierarchical structure, including low, middle and high level (Xuelong, 2015). The low-level classification methods mainly focus on describing subtle texture information and extract the original attributes of the scene pixel-by-pixel, such as SIFT, LBP, GIST, etc. The high-level methods are based on the low-level characteristics extracted to model attributes, containing a wealth of image semantic information, such as GoogleNet, AlexNet, CaffeNet, etc. But semantic gap always exists between low-level features and high-level semantic (Gu, 2016). In order to overcome the semantic gap and to solve the problem of overlapping definitions in hierarchical classification, the methods of scene modeling based on mid-level semantic features have drawn more and more attention. This paper selects a mid-

level scene method, Bag of Words (BoW) model. The BoW is originally used in text categorization, and the document is represented as a feature vector. For a text, the word order and grammar are ignored, and just regarded as a collection of some words, and each word in the text is independent. In short, we consider each document as a bag, and see what vocabulary is stored in this bag and then identify its category.

In this paper, SIFT features are detected and clustered as word lists of BoW. That is, BoW model is described with local descriptor(Zhou, 2013). At first, the SIFT feature extraction algorithm is used to extract feature points from each scene image as the visual words and each SIFT feature point contains 128 dimensions. Assuming the scene size is 96*96 pixels and divided into patches with size of 16*16 pixels with 50% overlapping between patches to make use of spatial information nearby, there would be 121 patches. If just one key point is extracted from each patch, there will be 121 key points in each scene. That is, each scene image is turned into a 121*128-dimension vector eventually.

Next, K-means algorithm is used to cluster SIFT points to build the vocabulary after setting the maximum number of iterations. And the number of clustering centers is selected from several hundred to thousand. Generally, the larger the amount of data is, the more clustering centers are. 100 centers are set for scene with size of 96*96 pixels here. After the clustering process is completed, 100 clustering centers are obtained as the vocabulary, each of which is a 1*128-dimensions vector.

And then, histogram of each scene, also known as word frequency, is computed based on the vocabulary obtained above. The initial value of histogram can be set to 0 and increment by 1. There are 121 SIFT points in each scene image. When the smallest-instance cluster center away from these points is determined, the value of the number between 1-100 represented by the nearest clustering center can be added 1 to form histogram. And finally the matrices of extracted BoW data can be obtained. Note that since the number of key points in each image is the same as 121, the normalization processing is not particularly critical. If not the same, the histogram must be normalized to prevent misclassification problems caused by the different numbers of extracted features. That is, we should divide the number of words by the total number of points to generate the word frequency.

In the next moment, how to match LBOs scene with BoW features is solved by the chosen classifier. There is study showing that the BoW method works better while combining with SVM classifier(Li, 2018). The selected Bow data by certain proportion obtained in the previous step is trained by the SVM classifier. For the new scene sample, the SIFT features are extracted first, and mapped into the dictionary to compute the histogram. Then the unknown scene can be predicted by the trained model using SVM classifier.

3. EXPERIMENTS

The experimental data sets used in this paper are the C-band and L-band full polarimetric SAR data acquired by AIRSAR in Tottori ken, Japan on Oct. 4, 2000 with 4.63–m azimuth resolution and 3.33–m range resolution, as shown in Figure 1. According to the corresponding high-resolution optical image, three LBO pixels (water, bare soil and shadow) can be labeled as ground truth.

(a) PauliRGB of C band   (b) PauliRGB of L band

■ water   ■ bare soil   ■ shadow
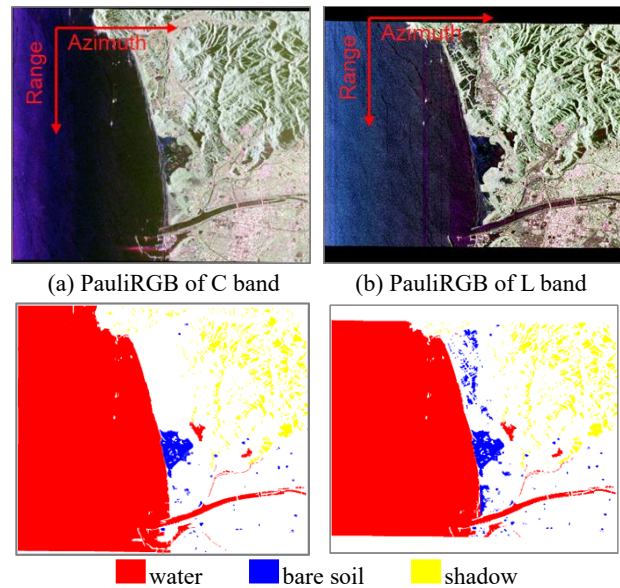(c) Ground truth of C band   (d) Ground truth of L band
Figure 1. The AIRSAR PauliRGB images and the corresponding ground truth.

The PauliRGB images are firstly segmented into different connected areas, of which the minimum enclosing squares can be determined, and areas with too small size are removed. We expand the length and width of each square at 50% and resize the expanded square to 96*96 pixels, and then label the resized squares as LBO scenes with reference to the ground truth. From the labeled scenes, 50% scenes of each LBO are selected randomly for training in SVM, as shown in Figure 2 and all scenes for test.
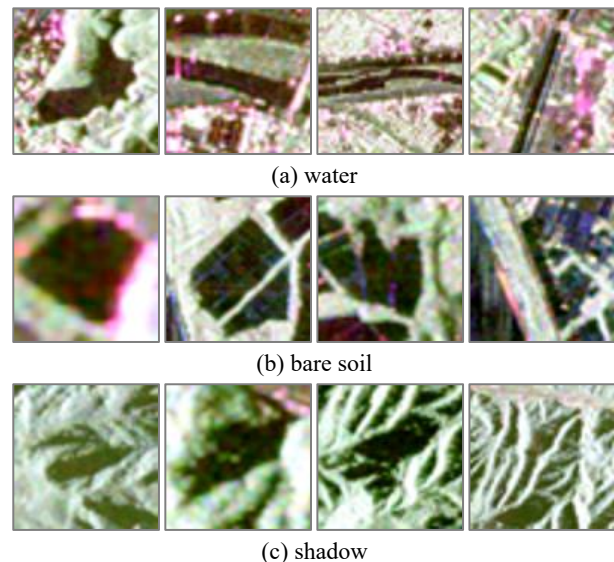

(a) water

(b) bare soil

(c) shadow

Figure 2. The training scenes selected for supervised classification.

Scenes with size of 96*96 pixels for training or test are divided into 16*16-pixels patches with 50% overlapping, from each of which a 128-dimension SIFT feature point is extracted. There are 121 patches in a scene image so each scene is turned into a 121*128-dimension vector actually. Then the SIFT points are clustered using K-means algorithm by setting 100 iterations and 100 cluster centers to build vocabulary. Histogram of each scene, also called word frequency, is computed in succession to obtain BoW data.

However, as a local description operator, SIFT feature can effectively resist affine transformation with rotation invariance and yet lacks some spatial relationship information among feature points. So that the three-layer spatial pyramid method is added to SIFT-based BoW, called Pyramid-BoW. The BOW and Pyramid-BOW is the same in the whole scene's statistics of word frequency, nevertheless, the main difference is the processing of word frequency. The former is in a global manner and the latter is hierarchical. At first layer, the Pyramid-BoW method divides one scene into 4*4 blocks to calculate the word frequency of 100 cluster centers and weight the statistics with 2^{-1}. And at second layer, the scene is divided into 2*2 blocks to carry on the word frequency statistics with 2^{-2} weight. And at the third layer, the histogram of the whole scene is computed with 2^{-2} weight. Finally, the three histograms are summarized to obtain the Pyramid-BoW data.

The 50% selected scenes are trained by SVM classifier using the Pyramid-BoW data and then all scenes are predicted by the trained model. To guarantee stabilization of the introduced algorithm, the selection of training scenes and classification steps are repeated 10 times to reach the average performance. The final classification results at pixel level are obtained after defining each connected area as some kind of LBO according to the predicted labels, as shown in Figure 3. After 10 times repetition, the overall accuracy (OA) is calculated as 94.1814% and Kappa coefficient is 0.8369. As for the LBOs, the water's classification accuracy is 0.92, the bare soil is 0.68, and the shadow is 0.99. The results in Figure 3 show that the three LBOs, water, bare soil and shadow, can be distinguished effectively, except that few bare soil scenes are misclassified as shadow in Figure 3(b), which is consistent with the quantitative analysis of classification accuracy. As a whole, the classification results are satisfactory.
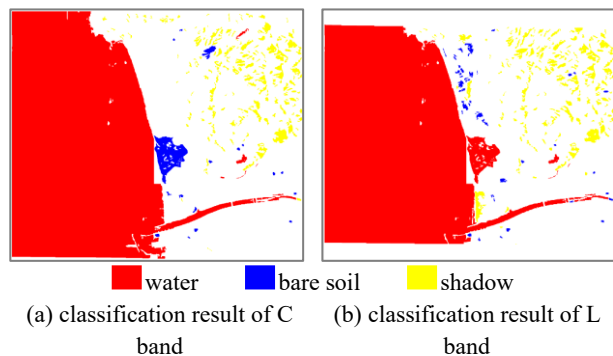

| ■ water | ■ bare soil | ■ shadow |

(a) classification result of C band
(b) classification result of L band

Figure 3. The LBOs classification results of scene method.

4. CONCLUSION

Because of the similar scattering mechanism of the typical LBOs, such as water, bare soil and shadow, it is difficult for the conventional classification algorithms based on scattering characteristics to distinguish them effectively. This paper introduced the Pyramid-BoW model, a scene classification method, and used SVM classifier for LBOs discrimination of PolSAR image, without relying on any polarimetric information. The experiments using AIRSAR data showed that the scene method could distinguish water, bare soil and shadow effectively with high overall accuracy and kappa of classification results.

The future work will focus on more tests on other data sets to evaluate LBOs-discrimination capacity of the scene model and make improvements to reduce misclassification between bare soil and shadow as far as possible.

REFERENCES

Cloude, S. R. and Pottier, E., 1997. An Entropy Based Classification Scheme for Land Applications of Polarimetric Sar. *IEEE Transactions on Geoscience and Remote Sensing* 35(1), pp. 68-78.

Gu, G. H., Han, X. Y., Chen, C. X. and Zhao, Y., 2016. Survey on Semantic Scene Classification Research. *Systems Engineering and Electronics* 38(4), pp. 936-948.

Lang, F., Yang, J., Zhao, L., Zhang, J. and Li, D., 2012. Polarimetric Sar Data Classification with Freeman Entropy and Anisotropy Analysis. *Acta Geodetica et Cartographica Sinica* 41(4), pp. 556-562.

Lee, J.-S. and Pottier, E., 2009. Polarimetric Radar Imaging: From Basics to Applications. Boca Raton, FL, USA, CRC press.

Lee, J., Grunes, M. R., Ainsworth, T. L., Du, L., Schuler, D. L. and Cloude, S. R., 1999. Unsupervised Classification Using Polarimetric Decomposition and the Complex Wishart Classifier. *IEEE Transactions on Geoscience and Remote Sensing* 37(5), pp. 2249-2258.

Li, L., Ling-Jun, T., Cheng-Yu, G., Liang, W. and Jun, T., 2018. Texture Classification: State-of-the-Art Methods and Prospects. *Acta Automatica Sinica* 44(X.), pp. X-X.

Liu, W., Yang, J., Zhao, J. and Yang, L., 2017. A Novel Method of Unsupervised Change Detection Using Multi-Temporal Polsar Images. *Remote Sensing* 9(11), pp. 1135-1154.

Pulvirenti, L., Marzano, F. S., Pierdicca, N., Mori, S. and Chini, M., 2014. Discrimination of Water Surfaces, Heavy Rainfall, and Wet Snow Using Cosmo-Skymed Observations of Severe Weather Events. *IEEE Transactions on Geoscience and Remote Sensing* 52(2), pp. 858-869.

Shi, L., 2012. The Low Backscattering Targets Classification in Urban Areas. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume I-7, pp 171-176.

Shi, L., Sun, W., Yang, J., Li, P. and Lu, L., 2015. Building Collapse Assessment by the Use of Postearthquake Chinese Vhr Airborne Sar. *IEEE Geoscience & Remote Sensing Letters* 12(10), pp. 2021-2025.

Xuelong, L. I., Shi, J. H., Dong, Y. S. and Tao, D. C., 2015. A Survey on Scene Image Classification. *Scientia Sinica* 45(7), pp. 827-848.

Zhao, J., Yang, J., Lu, Z., Li, P., Liu, W. and Yang, L., 2017. A Novel Method of Change Detection in Bi-Temporal Polsar Data Using a Joint-Classification Classifier Based on a Similarity Measure. *Remote Sensing* 9(8), pp. 846-862.

Zhao, L., Yang, J., Li, P., Lang, F. and Shi, L., 2013. Statistical Classification of Weak Backscattering Scatterers of Polsar Image. *Journal of Remote Sensing* 17(2), pp. 306-319.

Zhao, L., Yang, J., Li, P. and Zhang, L., 2014. Seasonal Inundation Monitoring and Vegetation Pattern Mapping of the Erguna Floodplain by Means of a Radarsat-2 Fully Polarimetric Time Series. *Remote Sensing of Environment* 152, pp. 426-440.

Zhou, L., Zhou, Z. and Hu, D., 2013. Scene Classification Using a Multi-Resolution Bag-of-Features Model. *Pattern Recognition* 46(1), pp. 424-433.