Ratio-Detector-Based Feature Extraction for Very High Resolution SAR Image Patch Indexing

Shiyong Cui, Corneliu Octavian Dumitru, and Mihai Datcu, *Fellow, IEEE*

Abstract—With the advent of very high resolution (VHR) synthetic aperture radar (SAR) images, local content description is becoming a critical issue for indexing. Conventional SAR image analysis techniques, like segmentation and pixel-level classification, are likely to fail as high-level semantic description should be considered for better discrimination. Therefore, we propose to use *image-patch-based* analysis method for SAR image interpretation. Inspired by ratio edge detector, in this letter, a new feature extraction method represented by the mean ratios in different directions is proposed for VHR SAR image content characterization. Based on the mean ratio, two simple yet powerful and robust features are proposed for SAR image patch indexing. One is the bag-of-word model using not only the basic statistics, i.e., local mean and variance, but also the mean ratios in different directions. The second one is an adaptation of the Weber local descriptor to SAR images by substituting the gradient with the ratio of mean differences in vertical and horizontal directions. To evaluate the proposed features, image patch indexing based on active learning using a SAR image database consisting of high-resolution TerraSAR-X patches is performed. Comparison with the state-of-the-art features, particularly texture features, has shown improved performance for SAR image categorization.

Index Terms—Bag-of-words (BoW), feature extraction, ratio detector, SAR image indexing, synthetic aperture radar (SAR), Weber local descriptor (WLD).

I. Introduction

RECENTLY, the advent of very high resolution (VHR) synthetic aperture radar (SAR) satellites has posed tough challenges in developing new features for SAR image indexing, particularly for structure features. The typical characteristic of VHR SAR images is that the local context consists of complex structure arrangements, resulting in much more scene classes. In this case, the conventional SAR image analysis techniques, like segmentation and pixel-level classification, are likely to fail as high-level semantic description of the local context should be considered for discrimination. Although texture features, such as gray-level co-occurrence matrix (GLCM) [1], Gabor filter [2], and Gaussian Markov random field (GMRF) [3], have played an important role in low- and medium-resolution SAR image interpretations, complex structure in a local context is becoming more critical for VHR SAR image characterization than texture features [4], which is demonstrated in Fig. 1. Textures

can well discriminate classes with fully developed speckle, like forest and agriculture. However, for urban scenes, there is an obvious structure configuration which is very important for semantic understanding and indexing.

Therefore, in this letter, a higher patch-level processing taking account of local structure is developed. In the literature, several works have been done for structure description. A multilevel local pattern histogram (MLPH) was proposed in [5], which shows better performance than GLCM, Gabor, and GMRF. To the best of our knowledge, the first work on patch-level indexing in remote sensing community was carried out in [6]. In [4], a patch-level contextual descriptor of SAR images based on Fourier transform was proposed and compared with GLCM, which has shown better performance in classification. Due to its ability in high-level image description, bag-of-word (BoW) model has been recently introduced to remote sensing community for image annotation [7], object classification [8], target detection [9], and land use classification [10]. In SAR community, only one work in [11] was done using BoW model, which is unfortunately a pixel-level classification.

In this letter, inspired by ratio edge detector, we propose to use the mean ratios in various directions for local context description of VHR SAR images. Based on the mean ratio, two simple yet powerful and robust features for high-resolution SAR image interpretation are proposed. One is the BoW model using not only the basic statistics, i.e., local mean and variance, but also the mean ratios in different directions. The second one is an adaptation of the Weber local descriptor (WLD) to SAR images by substituting the gradient with the ratio of mean differences (RMD) in vertical and horizontal directions. The contribution of this letter is that we propose to use the mean ratios for the description of complex structure in the local spatial context. Based on the mean ratio, two simple yet powerful feature extraction methods are proposed for SAR image patch indexing. The first one is an improvement of the BoW feature using not only the local mean and variance but also the mean ratios. The second one is an adaptation of WLD to SAR images.

This letter is organized as follows. In Section II, the ratio SAR edge detector is reviewed. The BoW feature and adapted WLD are presented in Sections III and IV, respectively. The test database, evaluation, and comparison are described in Section V. In the end, a conclusion is drawn in Section VI.

II. SAR Ratio Detector

Due to the presence of multiplicative speckle noise, conventional differential-based edge detectors are insufficient for SAR images. To combat against speckle, a ratio edge detector was proposed in [12]. The ratio detector is defined as the ratio of
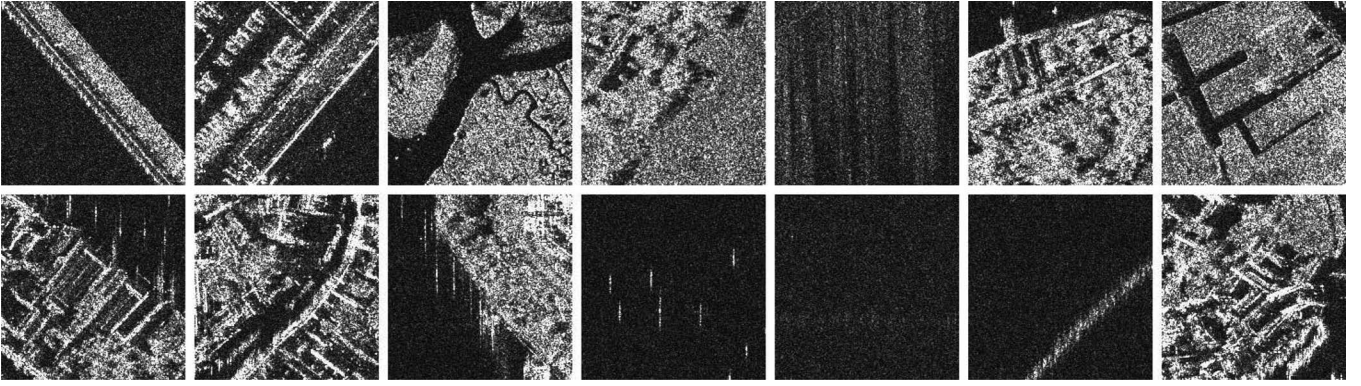
Fig. 1. All 14 classes and the semantic labels are bridge, harbor, river deposit, airport, breaking waves, vegetation and forest, agriculture, urban with water, urban, vegetation with water, buoy, water, water with boats, and vegetation with buildings. First row is from class 1 to 7, and second row is from class 8 to 14. In VHR SAR images, structure appears and becomes important for interpretation. Without consideration of complex structure arrangement in the local context, scene categorization cannot be efficiently achieved. Classical pixel-based classification and segmentation cannot solve this issue. Therefore, an image-patch-based analysis taking account of local structure is needed.
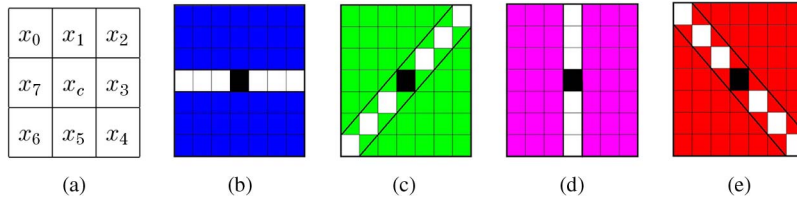

Fig. 2. Eight neighborhoods and four directions ($0°$, $45°$, $90°$, and $135°$) used for the definition of SAR image ratio detector.

the means of two neighborhoods on the opposite sides of the point. To detect all possible edges, the ratio detector should be applied in all possible directions. The edge response for a pixel is the maximum of all the edge responses in all directions. For the sake of computation, a local window centered at a pixel is split into two contiguous neighborhoods. Four principal directions, i.e., $0°$, $45°$, $90°$, and $135°$ are assumed in Fig. 2, and the ratios of local means (μ_1^i and μ_2^i), $i = 0, \dots, 3$, of the two neighborhoods are computed.

The edge responses of the four principal directions are defined as

$$r^i = 1 - \min\left(\frac{\mu_1^i}{\mu_2^i}, \frac{\mu_2^i}{\mu_1^i}\right), \qquad i = 0, \dots, 3. \qquad (1)$$

The final edge response for a pixel is given as $R = \max(r^i)$, $i = 0, \dots, 3$, which is referred as the mean ratio in this letter. The mean ratios in all the assumed directions are used as low-level features to improve the BoW feature in the next section.

Based on the ratio detector, RMD is defined as

$$R_{\text{SAR}} = \frac{D_v}{D_h} = \frac{\mu_1^1 - \mu_2^1}{\mu_1^3 - \mu_2^3} \qquad (2)$$

where $D_v = \mu_1^1 - \mu_2^1$ and $D_h = \mu_1^3 - \mu_2^3$ are the differences of the local means of the two neighborhoods corresponding to both vertical and horizontal directions, the two blue and magenta regions shown in Fig. 2(b) and (d). The RMD can be considered as an adaptation of the gradient for SAR images. Note that the RMD is defined based on a local window. Therefore, we can define a multiscale RMD by a series of increasing windows, i.e., $s = 3, 5, 7, \dots$. The RMD is used as an alternative to gradient in the adaptation of WLD to SAR images in the next section.

III. BoW Model

In the BoW model, an image is represented as a collection of local features, i.e., key points detected by scale-invariant feature transform (SIFT) [13] or densely sampled patches. Local features are extracted from the local neighborhoods around SIFT points or densely sampled patches. To construct the vocabulary, a clustering, usually k-means, is performed to find clusters. Cluster centers are used as vocabulary for computing word occurrence histogram. After vocabulary generation, each local feature is assigned usually to the nearest cluster, and thus, an image can be represented as the word occurrence histogram. Recent study in [14] tries to assign a local feature to multiple words. For the sake of simplicity, nearest word assignment is employed in this letter. One critical step in this framework is local feature extraction, which has attracted increasing interests in developing local descriptor in computer vision community. However, in SAR community, almost no solution has been proposed to develop SAR image local descriptor, which is the main motivation of this letter. We started simply from local statistics, mean and variance, of the original patches, which is abbreviated as BoW_MV in the following. For fully developed speckle, the relation between mean and variance given as $L = m^2/\sigma$ holds, where L is the number of looks, m is the mean, and σ is the variance. However, for the areas of strong structures, this relation does not hold any more. Therefore, we propose to incorporate the mean ratios in different directions in addition to the local mean and variance, in the process of vocabulary generation. In this case, the feature is abbreviated as BoW_MVR in the following sections. The BoW_MVR features of five classes are shown in Fig. 3. As can be seen, the word histogram has quite strong discriminative ability for SAR image categorization. Quantitative evaluation is conducted in Section V.
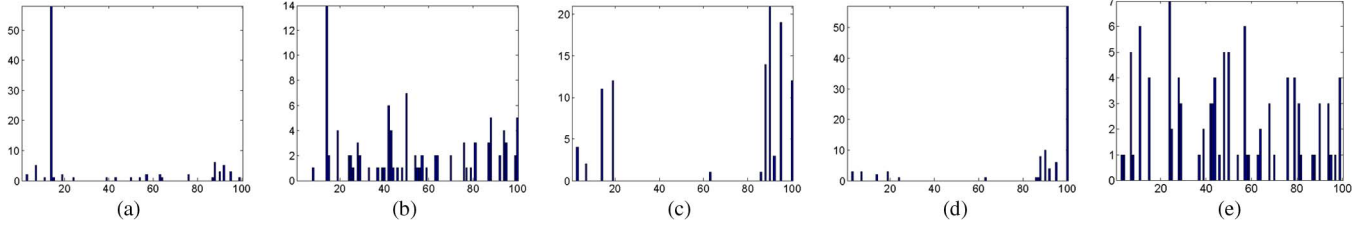
Fig. 3. (From left to right) Word histograms of the first (bridge), second (harbor), third (river deposit), fourth (vegetation and forest), and ninth (urban) classes after incorporating the mean ratios.
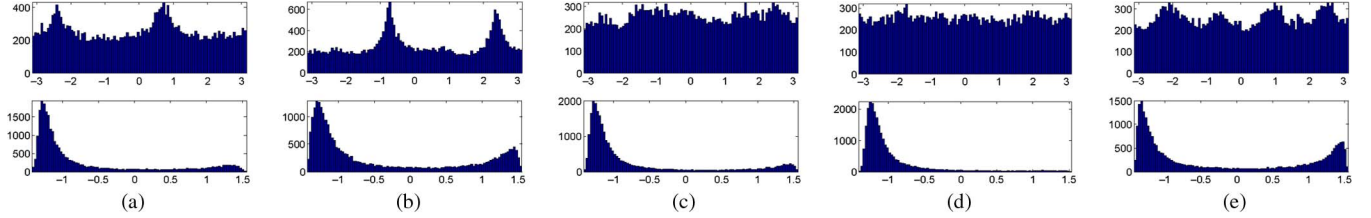

Fig. 4. (From left to right) Histograms of orientation and differential excitation of the first (bridge), second (harbor), third (river deposit), fourth (vegetation and forest), and ninth (urban) classes.

IV. ADAPTATION OF WLD TO SAR IMAGES

WLD was proposed in [15] for optical image retrieval by the inspiration of Weber's law, which states that the change of a stimulus that will be just noticeable is a constant ratio of the original stimulus. If the change is smaller than this constant ratio, it cannot be recognized. Based on this idea, WLD is proposed for texture characterization, which is composed of two components: differential excitation ξ and orientation θ.

The differential excitation is defined as

$$\xi(x_c) = \arctan\left[\sum_{i=0}^{n-1}\left(\frac{x_i - x_c}{x_c}\right)\right] \quad (3)$$

where n is the number of neighboring pixels, which is eight in the case of Fig. 2(a). The orientation is given as the gradient orientation, defined by

$$\theta(x_c) = \arctan 2\left(\frac{x_7 - x_3}{x_5 - x_1}\right). \quad (4)$$

Based on the two terms, a joint histogram can be constructed, followed by converting to a 1-D histogram, which is the WLD descriptor. The advantage of this feature is that it considers not only the local contrast but also the structure information represented by gradient histogram.

Although Weber's law is a theory from visual perception and there is no visual aspect for SAR images, the principle for feature extraction is applicable to SAR images. Nevertheless, in the case of SAR images, multiplicative speckle noise decreases dramatically its discriminative ability for image indexing. To combat against speckle, we propose a solution to replace the gradient in WLD by the RMD, leading to the following orientation component

$$\theta(x_c)_{\text{SAR}} = \arctan 2(R_{\text{SAR}}) = \arctan 2\left(\frac{D_v}{D_h}\right). \quad (5)$$

Furthermore, the differential excitation is adapted as

$$\xi(x_c)_{\text{SAR}} = \arctan\left[\sum_{i=0}^{3}\sum_{j=1}^{2}\left(\frac{\mu_j^i - x_c}{x_c}\right)\right]. \quad (6)$$

As an example, the histograms of the orientation and differential excitation for SAR images from five classes are shown in Fig. 4. As can be seen, the adapted WLD is discriminative for structure. Visually, the five classes can be well discriminated.

Based on the two adapted components $\theta(x_c)_{\text{SAR}}$ and $\xi(x_c)_{\text{SAR}}$, WLD for SAR images can be defined as the joint histogram $H(\xi(x_c)_{\text{SAR}}, \theta(x_c)_{\text{SAR}})$. Therefore, there are two adjustable parameters, the numbers of bins for excitation C and orientation T. Following the same strategy as WLD, this joint histogram is converted to a 1-D histogram. The adapted WLD includes not only local statistics but also local structure information, resulting in an improved performance in SAR image indexing.

V. EVALUATION AND DISCURSION

A. Image Test Database

To evaluate the performances of the two proposed features, BoW_MVR and the adapted WLD, a SAR image database was prepared by tiling one detected TerraSAR-X image of 10881×15782 pixels covering the Venice (Italy) area. The data format is multilook ground-detected spatially enhanced high-resolution spotlight image with single polarization (HH) and 2.9-m resolution. The pixel spacing is 1.25, and the incidence angle is $38°$. The numbers of looks in range and azimuth are 3.08 and 2.60, respectively. The TerraSAR-X image was tiled into patches of 160×160 covering a 200×200 m on the ground. From about 1026 patches, 17 different classes were extracted. From these classes, we selected 14 classes shown in Fig. 1 with more than ten example patches of each.
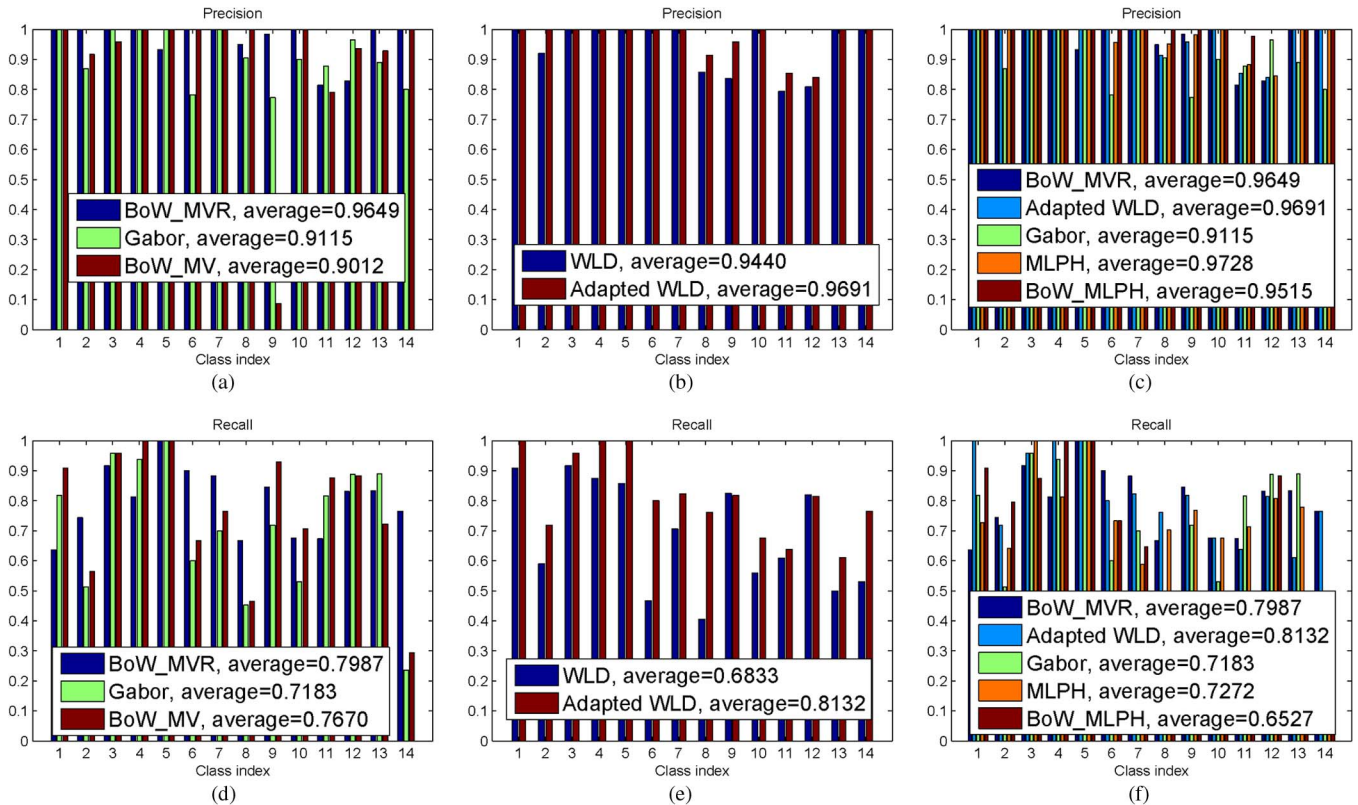
Fig. 5.   Performance comparison of different features for SAR image indexing. (a) and (d) Precision and recall comparison of BoW and Gabor texture. (b) and (e) Precision and recall comparison of two BoW models. (c) and (f) Precision and recall comparison of WLD and adapted WLD.

B. Evaluation and Results

All experiments conducted hereinafter are based on an active learning system for interactive SAR image indexing. Two important components in our active learning system are two modules for classifier training using the labeled images and sample selection which selects the most informative samples for manual labeling. These two components work alternatively, which can significantly reduce human labeling effort and achieve better performance for image indexing. The classifier we adopted is the C-support vector machine (SVM), and the kernel is a chi-square function that is effective and efficient for histogram-based feature. The only parameter in C-SVM is the penalty parameter C. This parameter is set to 1000 empirically. Samples close to the class boundary are selected in each iteration from the unlabeled pool for manual labeling, which are the most important ones to train a classifier effectively and quickly.

Three experiments have been carried out in this section for demonstration. The first one is to demonstrate the advantage of BoW_MV over Gabor texture and the improvement of BoW_MVR, while the second one is to compare WLD with the adapted WLD. The last evaluation is to compare these two features with the state-of-the-art SAR features. Furthermore, BoW feature using MLPH as low-level feature, abbreviated as BoW_MLPH, is also generated and compared because BoW feature is a midlevel feature, which can be generated using any low-level features. The Gabor texture features used in all the three experiments are the mean and variance of each subband [2]. Performances in all these experiments are measured by precision and recall.

In the BoW setting, each image is tiled into patches with size 16×16 pixels, corresponding to the window size for local feature extraction. In the first test, we compare BoW_MV with Gabor texture features generated with four scales and six orientations. The clustering algorithm we adopted for vocabulary generation is enhanced Linde-Buzo-Gray algorithm [16], which converges toward a better optimum compared with k-means. The number of clusters we assumed is 100 as seeking for an appropriate number of clusters is out of the scope of this letter. It has been demonstrated that there is no big improvement using large vocabulary size in [8]. Each feature vector is assigned to its closest word in the feature space. The precision and recall are shown in Fig. 5(a) and (d). As can be seen surprisingly from the results, BoW_MV performs better than Gabor texture feature except for only two (12th and 13th) classes. These two classes are pure water and water with boat, respectively, which are quite homogeneous as shown in Fig. 1. The average recalls of these two features are 71.83% and 76.70%. Five percent improvement in recall is achieved using only the local mean and variance, which is a quite promising improvement. The precision and recall of both BoW_MVR and BoW_MV are also shown in Fig. 5(a) and (d). The performance of BoW_MVR is slightly better than that of BoW_MV. The average precision and recall improve by 6.37% and 3.17%, respectively. However, for well-structured urban classes, particularly the ninth class, there is a big improvement in recall.

The second evaluation is conducted to compare the adapted WLD with WLD, where the features are extracted from the entire patch without tiling into subpatches. In both procedures of feature extraction, the involved parameters (the number of

bins for orientation and excitation) are set to the same values, $C = 18$ and $T = 8$. The window sizes used in local feature extraction for the adapted WLD and WLD are seven and three, respectively. A larger window size is needed for SAR images to alleviate the impact of speckle. The precision and recall are shown in Fig. 5(b) and (e). It can be clearly seen that both precision and recall are significantly improved for all classes, which confirms the effectiveness and robustness of the adapted WLD to SAR images. The average improvements of precision and recall are 2.51% and 12.99%, respectively. The improvement of the adapted WLD can be well explained by the fact that SAR images are impacted significantly by multiplicative noise, which is not the case for optical images. The RMD alleviates the influence of multiplicative speckles by averaging over all pixels in the neighborhoods and taking the ratio of means. The effectiveness of RMD for structure description in SAR images is also confirmed by the accuracy improvement.

The last experiment is carried out to compare with state-of-the-art features and also the BoW feature based on these features. The comparison of the state-of-the-art features was done in our previous work [17]. In that work, we compared GLCM, Nonlinear short time Fourier transform (NSTFT), Gabor texture, and quadrature mirror filter (QMF) texture features with different parameter setting. We concluded in that paper that Gabor texture feature performs better compared with the others. As shown in the first experiment, Gabor feature is worse than BoW_MV. Therefore, in the following, we do not consider GLCM, NSTFT, QMF, and Gabor features and compare only with MLPH and related BoW feature BoW_MLPH. In the parameter setting of MLPH, there are mainly two parameters which are the number of levels and the increase rate of the threshold. They are set to five and three, respectively. Different from the original implementation in [5] where a sliding window is applied at each pixel to extract a local pattern histogram, we applied the thresholding directly to the patch and calculated a local pattern histogram, which means that the patch of 16×16 is equivalent to the sliding window in the original implementation. The precision and recall of these four features are shown in Fig. 5(c) and (f). As a baseline for comparison, precision and recall of Gabor features are also plotted. MLPH is slightly better than Gabor. Similarly, BoW_MLPH is worse than MLPH. From this result, we see that there is not always gain to apply the BoW model to a low-level feature. In contrast, as can be seen, adapted WLD and BoW_MVR achieve comparable performance with similar precisions of 96.91% and 96.49% and recalls of 81.32% and 79.87%, because both of them consider not only local statistics but also local structure information in the spatial context.

VI. CONCLUSION

To overcome the drawbacks of pixel-level VHR SAR image analysis techniques, an image-patch-based method has been proposed for SAR image interpretation. Inspired by ratio edge detector, in this letter, a new feature extraction method represented by the mean ratios in different directions is proposed for VHR SAR image characterization. Based on the mean ratio, two simple yet powerful and robust features are proposed for SAR image patch indexing. One is the BoW model using not only the basic statistics, i.e., local mean and variance, but also the mean ratios in different directions. The other is an

adaptation of WLD to SAR images by substituting the gradient with the ratio of mean differences in vertical and horizontal directions. These two features are evaluated and compared with state-of-the-art features based on an active learning system using a database consisting of 1000 TerraSAR-X images. The evaluation and comparison have demonstrated the effectiveness of the proposed features in structure description because both of them consider not only the local statistics but also the local structure context. The adaptation of WLD shows significant improvement compared with WLD in SAR image indexing. In addition, BoW model using only the mean and variance as low-level features has strong discriminative ability and is confirmed, which shows better discriminative ability after incorporating the mean ratio.

As a future work, multiscale patch is considered to enhance the discriminative ability of the features proposed in this letter.

REFERENCES

[1] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-3, no. 6, pp. 610–621, Nov. 1973.

[2] B. S. Majunath and W. Y. Ma, "Texture features for browsing and retrieval of image data," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 8, pp. 837–842, Aug. 1996.

[3] D. A. Clausi and B. Yue, "Comparing cooccurrence probabilities and Markov random fields for texture analysis of SAR sea ice imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 1, pp. 215–228, Jan. 2004.

[4] A. A. Popescu, I. Gavat, and M. Datcu, "Contextual descriptors for scene classes in very high resolution SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 1, pp. 80–84, Jan. 2012.

[5] D. Dai, W. Yang, and H. Sun, "Multilevel local pattern histogram for SAR image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 2, pp. 225–229, Mar. 2011.

[6] C.-R. Shyu, M. Klaric, G. J. Scott, A. S. Barb, C. H. Davis, and K. Palaniappan, "GeoIRIS: Geospatial information retrieval and indexing system—Content mining, semantics modeling, and complex queries," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 4, pp. 839–852, Apr. 2007.

[7] 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, Jan. 2010.

[8] S. Xu, T. Fang, D. Li, and S. Wang, "Object classification of aerial images with bag-of-visual words," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 2, pp. 366–370, Apr. 2010.

[9] H. Sun, X. Sun, H. Wang, Y. Li, and X. Li, "Automatic target detection in high-resolution remote sensing images using spatial sparse coding bag-of-words model," *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 1, pp. 109–113, Jan. 2012.

[10] Y. Yang and S. Newsam, "Bag-of-visual-words and spatial extensions for land-use classification," in *Proc. 18th SIGSPATIAL Int. Conf. Adv. GIS*, New York, 2010, pp. 270–279.

[11] J. Feng, C. L. Jiao, X. Zhang, and D. Yang, "Bag-of-visual-words based on clonal selection algorithm for SAR image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 4, pp. 691–695, Jul. 2011.

[12] R. Touzi, A. Lopes, and P. Bousquet, "A statistical and geometrical edge detector for SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 26, no. 6, pp. 764–773, Nov. 1988.

[13] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004.

[14] J. C. van Gemert, C. J. Veenman, A. W. M. Smeulders, and J.-M. Geusebroek, "Visual word ambiguity," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 7, pp. 1271–1283, Jul. 2010.

[15] J. Chen, S. Shan, C. He, G. Zhao, M. Pietikainen, X. Chen, and W. Gao, "WLD: A robust local image descriptor," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 9, pp. 1705–1720, Sep. 2010.

[16] G. Patané and M. Russo, "The enhanced LBG algorithm," *Neural Netw.*, vol. 14, no. 9, pp. 1219–1237, Nov. 2001.

[17] J. Singh, C. O. Dumitru, and M. Datcu, "Selection of relevant features and TerraSAR-X products for classification of high resolution SAR images," in *Proc. EUSAR*, Nuremberg, Germany, Apr. 23–26, 2012, pp. 243–246.