In: Wagner W., Székely, B. (eds.): ISPRS TC VII Symposium – 100 Years ISPRS, Vienna, Austria, July 5–7, 2010, IAPRS, Vol. XXXVIII, Part 7B

WEAKLY SUPERVISED POLARIMETRIC SAR IMAGE CLASSIFICATION WITH MULTI-MODAL MARKOV ASPECT MODEL

Wen Yang, Dengxin Dai, Jun Wu, Chu He

School of Electronic Information, Wuhan University
LuoJia Hill, Wuhan 430072, China
yangwen@whu.edu.cn
http://dsp.whu.edu.cn

Commission VII

KEY WORDS: Land Cover, Classification, Polarization, SAR, RADARSAT, Imagery

ABSTRACT:

In this paper, we present a weakly supervised classification method for a large polarimetric SAR (PolSAR) imagery using multi-modal markov aspect model (MMAM). Given a training set of subimages with the corresponding semantic concepts defined by the user, learning is based on markov aspect model which captures spatial coherence and thematic coherence. Classification experiments on RadarSat-2 PolSAR data of Flevoland in Netherlands show that this approach improves region discrimination and produces satisfactory results. Furthermore, multiple diverse features can be efficiently combined with multi-modal aspect model to further improve the classification accuracy.

1 INTRODUCTION

During the last decade, several space-borne sensors with polarimetric SAR (PolSAR) imaging have been lauched and produce terrabytes of PolSAR images. PolSAR remote sensing offers an efficient and reliable means of collecting information required to extract geophysical and biophysical parameters from Earth's surface, which shows potential for improved results in many successful applications. As it becomes increasingly viable to acquire, store, order and share large amounts of polarimetric SAR data, accurate and ease-to-use supervised classification method is crucial to extracting information from these datasets.

Earlier supervised classification methods for polarimetric SAR data are mainly pixel-based schemes. The widely used methods are the maximum likelihood classification based on the complex Wishart distribution (Lee et al., 1994) and its variations (Lee et al., 2001) (Beaulieu and Touzi, 2004). The classification performances of these methods are affected by speckle seriously since they are unable to capture and utilize the spatial information in the scene. To overcome this problem, region-based methods have been employed, which use a over-segmentation step (or grid partition step) and form groups of pixels that represent homogeneous regions. In (Wu et al., 2008), wu et al. proposed a region-based classification method for polarimetric SAR images with a Wishart Markov Random Field model, which can efficiently use the statistical properties of the data and the spatial relation of neighboring pixels. Ersahin et al. (Ersahin et al., 2010) proposed to use spectral graph partitioning approach for segmentation and classification of POLSAR data,and achieved promising classification accuarcy superior to the Wishart classifier. Recently, classifiers originated from machine learning and pattern recognition domain have attracted more attention, such as neural networks (Shimoni et al., 2009), support vector machine(SVM) (Lardeux et al., 2009), and Random Forests (Zou et al., 2010). These methods are also usually implemented on region level, and they can easily handle many sophistical image features and get remarkable performance. However, existing supervised classification methods most require the labor-intensive and time-consuming works to label every pixel in the training samples. Furthermore, as to general user, it is very difficult to make ground truth for pixel-level labeled training samples in SAR image, sometimes only experts of SAR image interpretation are qualified for this job. In this work, we are interested in weakly supervised classification of PolSAR images, which is aimed at partitioning a PolSAR scene into their constituent semantic-level regions with only keywords labeled training data.

In this study, we present a solution using a multi-modal markov aspect model proposed by Verbeek and Triggs (Verbeek and Triggs, 2007), which can be learned from image-level keywords without detailed pixel-level labeling. The whole classification process consists of four cascaded stages. In the first stage, we partition the whole PolSAR scene into hundreds of subimages. From each subimage we extract overlapping patches on a grid, representing them by polarimetry, intensity and texture descriptors. We assume that each subimage patch belongs either to one of the label classes or to a vague background class "void". Then, we model each subimage as a mixture of latent aspects with a multi-modal markov aspect model which can be learnt from image-level keywords. Next, we use an efficient expect maximization (EM) algorithm to learn the model and apply loopy belief propagation (LBP) (Yedidia et al., 2005) inferring algorithm to label every patch in the test subimages with the trained model. Finally, we apply a over-segmentation based soft mapping to propagate patch-level labeling to pixel-level classification, and group the large PolSAR scene classification result from the labeled subimages.

The rest of the paper is organized as follows. Section 2 briefly introduces the three type features we used for classifying PolSAR images. Section 3 describes the baseline classifier-Wishart maximum likelihood classifier. The multi-modal markov aspect model is reviewed in Section 4. Section 5 gives comparative experimental results and quantitative evaluation, and section 6 concludes the paper.

2 FEATURE DESCRIPTORS FOR POLSAR IMAGES

Classification problem is challenging because the instances in SAR images belonging to the same class usually have very high intraclass variability. To overcome this problem, one strategy is to design feature descriptors which are highly invariant to the variations present within the classes, however none of the feature descriptors will have the same discriminative power for all classes. The other widely accepted strategy is that, instead of using a single feature type for all classes, it is better to combine multiple diverse and complementary features based on different aspects. Therefore, we extract multiple polarimetric and low-level image features for describing the small patches in each PolSAR subimage. A more detailed description of these feature parameters is given below:

2.1 Polarimetry

PolSAR is sensitive to the orientation and characters of target and thus yields many new polarimetric signatures which produce a more informative description of the scattering behavior of the imaging area. There are many polarimetric descriptors summarized in (Shimoni et al., 2009). For simplicity, we just use the nine parameters obtained by Huynen decomposition (Huynen, 1990).

Given a scattering matrix measured in the orthogonal linear (h, v) basis, the classical 2×2 Sinclair scattering matrix S can be obtained through the construction of system vectors.

$$S = \begin{pmatrix} S_{HH} & S_{HV} \\ S_{VH} & S_{VV} \end{pmatrix} \tag{1}$$

The coherency matrix is constructed from a scattering vector in the base of Pauli basis. In the monostatic backscattering case, for a reciprocal target matrix, the reciprocity constrains the Sinclair scattering matrix to be symmetrical, that is, $S_{HV} = S_{VH}$. Thus, the target vectors k_p can be constructed based on the Pauli basis sets, respectively. With this vectorization we can then generate the coherency matrix T as follows,

$$k_p = \frac{1}{\sqrt{2}} \begin{bmatrix} S_{HH} + S_{VV} \\ S_{HH} - S_{VV} \\ 2S_{HV} \end{bmatrix}, [T] = \left\langle k_p \cdot k_p^{*T} \right\rangle, \tag{2}$$

The Huynen decomposition (Huynen, 1990) is the first attempt to use decomposition theorems for analyzing distributed scatters. In the case of coherence matrix, this parametrization is

$$[T] = \begin{bmatrix} 2A_0 & C - jD & H + jG \\ C + jD & B_0 + B & E + jF \\ H - jG & E - jF & B_0 - B \end{bmatrix} \tag{3}$$

The set of nine independent parameters of this parametrization allows a physical interpretation of the target: A_0, B_0+B, B_0-B, C, D, E, F, G, H. The nine Huynen parameters are useful for general target analysis without reference to any model, and each of them contains real physical target information.

2.2 Texture

The Gray Level Co-occurrence Matrix (GLCM), Gabor filters, Gaussian Markov random fields (GMRF) Texture are three widely used features for SAR image texture segmentation. Former experiments show that GMRF yields the best performance in terms of classification accuracy, although it has high computational complexity in high order case (Clausi, 2001). The Gaussian Markov Random Field models characterize the statistical dependency between a pixel and its neighbors by representing the gray level intensity at site s, as a linear combination of gray levels in a neighborhood set $N(s)$ around s, and Gaussian zero mean stationary noise. The specific definition of neighbours and their influence on other points give GMRF the freedom to model many types of

5	4	3	4	5
4	2	1	2	4
3	1	o	1	3
4	2	1	2	4
5	4	3	4	5

Figure 1: The first to fifth-order MRF neighborhood system

textures. GMRF model is fitted on the patches, and the model parameters are used to form a multi-dimensional feature space. Figure 1 illustrates the different orders of MRF, which is related to the location of neighbors. To balance the computation complexity and classification accuracy, we use a four-order GMRF, which can be expressed as follows:

$$f(m, n) = \sum_{(t,s)\in N} \theta(t, s) f(m - t, n - s) + e(m, n) \tag{4}$$

where N represents the 20 neighborhoods, and $e(m, n) \sim N(0, \sigma^2)$ with zero mean and variance σ^2. For each pixel, we train the mean μ, σ, and due to the symmetry of correlation function, only the 10 parameters,$\theta(t, s)$, $(t, s) \in N$,over a window W centered on this, by Least Square Estimation (LSE). The feature space is formed as 12 dimensions vector $\mu, \sigma, \theta(t, s), (t, s) \in N$.

2.3 Intensity

In our previous work, we learn that histogram is a simple but informative descriptor for single polarimetric SAR imagery classification. For full polarimetric SAR images, unlike the SPAN histogram used in (He et al., 2008), we propose to use the multichannel histogram which is a cumulative enumeration of its underlying HH, HV and VV channels.

3 BASELINE CLASSIFIER-SUPERVISED ML CLASSIFICATION OF [T] MATRIX DATA

According to Bayes optimal decision rule, a pixel is assigned to the most probable class conditionally to the observation over the pixel under consideration. If the prior probabilities are supposed to be equal, the optimal decision rule reduce to the maximum likelihood supervised segmentation.

It has been shown that a n-look coherence matrix follows a complex Wishart distribution with n degrees of freedom, $W_c(n, [\sum])$, given by

$$\begin{aligned} P([T]) \quad &= \quad \frac{n^{qn}|[T]|^{n-q} exp(-tr(n[\Sigma]^{-1}[T]))}{K(n,q)|[\Sigma]|^n} \\ &with \quad K(n,q) = \pi^{q(q-1)/2 \prod_{i=1}^{q} \Gamma(n-i+1)} \end{aligned} \tag{5}$$

where q stands for the number of elements of target vector, equal to 3 in the monostatic case, $|\cdot|$ represents the determinant, tr represents the trace of a matrix and $\Gamma(\cdot)$ denotes the Gamma function. A pixel p can be assigned to a class $\{\Theta_i, \dots, \Theta_M\}$ in maximum likelihood way, according to the following steps,

- Initialize pixel distribution over M classes from training samples;
- For each class, $[\hat{\Sigma}_i = \frac{1}{N_i} \sum [T] \in \Theta_i]$, where $[\Sigma_i]$ is the coherence matrix of class Θ_i computed during the training phase;

- For each pixel, $[T] \in \Theta_i$ if $d([T], \Theta_i) < d([T]/\Theta_j), j = 1, \ldots, M, j \neq i$

In the following experiments, we will employ the pixel-based and patch-based wishart ML classifier as the baselines.

4 WEAKLY SUPERVISED CLASSIFIER-MARKOV ASPECT MODEL

Recently, many research works on labeling natural images focus on the utilization of high-level semantic representation with topic models, such as the Probabilistic Latent Semantic Analysis (PLSA) (Hofmann, 2001) or its bayesian form, the Latent Dirichlet Allocation (LDA) (Blei et al., 2003). They consider visual words as generated from latent aspects (or topics) and express image as combination of specific distributions of topics, which can solve some cases of visual "polysemy" by capturing thematic coherence (image-wide correlations). Verbeek and Triggs (Verbeek and Triggs, 2007) proposed to use markov aspect model that captures both spatial coherence (local correlations between labels) and thematic coherence (image-wide correlations), and further employed a multi-modal aspect model to combine multiple cues for improving classification accuracies. Experimental results on the Microsoft Research Cambridge data sets show their model significantly improves the region-level classification accuracy. Li et al. (Li and Perona, 2005) proposed two variations of LDA to generate the intermediate theme representation to learn and recognize natural scene categories, and reported satisfactory categorization performances on a large set of complex scenes.

In remote sensing domain, Liénou et al. (Liénou et al., 2010) proposed to exploit the LDA model to semantically annotate panchromatic QuickBird images with 60-cm resolution, and demonstrated that using simple features such as mean and standard deviation for the LDA-image representation can lead to satisfying labeling results. However, PLSA is computationally more efficient than LDA and it has comparable accuracy in practice (Verbeek and Triggs, 2007). In this work, we use multi-modal PLSA-MRF framework (Verbeek and Triggs, 2007) for polarimetric SAR image classification, which naturally introduces the spatial information by combining Markov Random Fields with Probabilistic Latent Topic Models. PLSA-MRF basically just adds pairwise MRF couplings to the PLSA label inference process, we use LBP for MRF inference and EM algorithm for maximum likelihood estimation of the model as done in (Verbeek and Triggs, 2007). Unlike multi-modal LDA (Li and Perona, 2005), multi-modal PLSA-MRF (i.e. multi-modal markov aspect model) assumes that the different types of features are independent given the topic.

$$P(w|d) = \sum_{t=1}^{T} P(w^{huy}|t)P(w^{hist}|t)P(w^{gmrf}|t)P(t|d) \quad (6)$$

$$P(t|w,d) = \frac{P(w^{huy}|t)P(w^{hist}|t)P(w^{gmrf}|t)P(t|d)}{P(w|d)} \quad (7)$$

5 RESULTS AND DISCUSSION

We validate the above-mentioned multi-modal markov aspect model on the semantic annotation of a large scene Radarsat-2 Polarimetric SAR image. In this Section, we present our experimental setup, show a detailed performance evaluation illustrated with the classification results, and we finally discuss the limits of the labeling algorithm.

5.1 Data Description and Experimental Setup

The experiments are performed on RadarSat-2 fully polarimetric SAR images of Flevoland in Netherlands, with 12m $\times$8m resolution at fine quad-pol mode. The PolSAR scene to be labeled is of size 4000 $\times$ 2400 pixels, which mainly contains four classes: woodland(Wo), cropland(Cr), water(Wa), building area(Bu). We divide the PolSAR scene into 240 subimages, each subimage is 200 $\times$200 pixels. Fig.2 shows 8 examples of keywords labeled training subimages, and we use 40 subimages with such keywords annotation as the training set.

5.2 Post-processing with over-segmentation mapping

In our labeling algorithms, learning and inference take place at the patch level, but the results are propagated to pixel level for visualization and performance quantification. We apply a mean-shift based over-segmentation mapping to map the patch-level labelings to pixel-level classification. It combines the nearest mapping result (we compute the class label at the pixel level as the nearest patch label) with a low level over-segmentation since segment boundaries can be expected to coincide with the image edges, which can reduce the block effect of the nearest mapping and also can improve the accuracy slightly. Here we compute the over-segmentation with the Edge Detection and Image Segmentation (EDISON) System of Mean Shift (Comaniciu and Meer, 2002) implementation. The parameters of the segmentation are chosen to mostly over-segment the subimages.

5.3 Classification results and quantitative evaluation

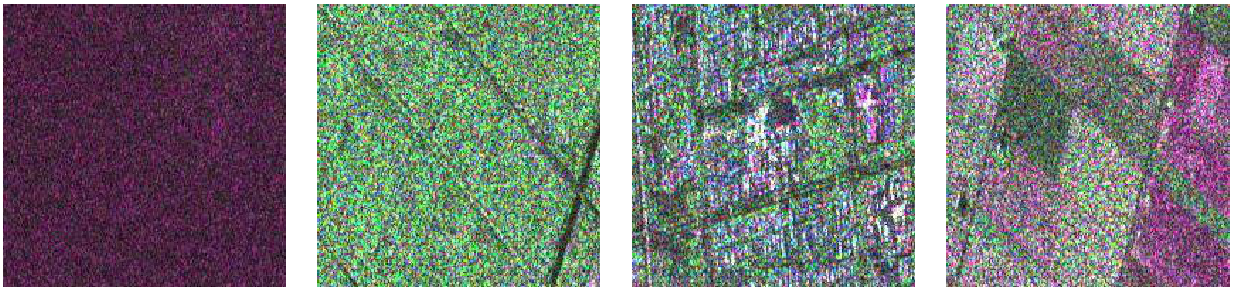
For quantitative evaluation of the classification accuracy with different features and classifiers, we select a region from the original test site of size 1400 $\times$ 1200 with corresponding elaborately labeled ground truth. Pixels are assigned to four semantic classes or void. The four classes are building area, woodland, water and cropland. The void pixels either do not belong to one of the four classes or lie near boundaries between classes and were labeled as void to simplify the task of manual segmentation. About 6% of the pixels are unlabeled ("void) in the evaluation data. Table 1 gives the performance of different classifiers and features. We can find that multi-modal markov aspect model using image-level labeled training data outperforms traditional Wishart ML methods with detailed pixel-level labeled training data by %3.8. In fact, even when using only one feature-GMRF or Hist, they provide pixel-level classification accuracies outperform those of Wishart ML classifier trained using detailed pixel-level labelings by 0.5% or 2.1%, respectively. The classification results of the original test site (4000 $\times$ 2400) using multi-modal markov aspect model and Wishart ML (pixel-based and patch based) are presented in Fig.4.

Est. \ True	Wo	Wa	Bu	Cr	ave.acc.
ML-pixel	86.2	90.5	34.1	76.3	**71.3**
ML-patch	91.5	73.8	44.3	78.1	**72.8**
MMAM-Huy	75.1	70.0	80.0	57.6	**69.4**
MMAM-GMRF	89.3	64.3	86.6	52.4	**71.8**
MMAM-Hist	78.1	81.3	94.0	51.9	**73.4**
MMAM-All	81.8	77.0	90.0	59.1	**75.1**

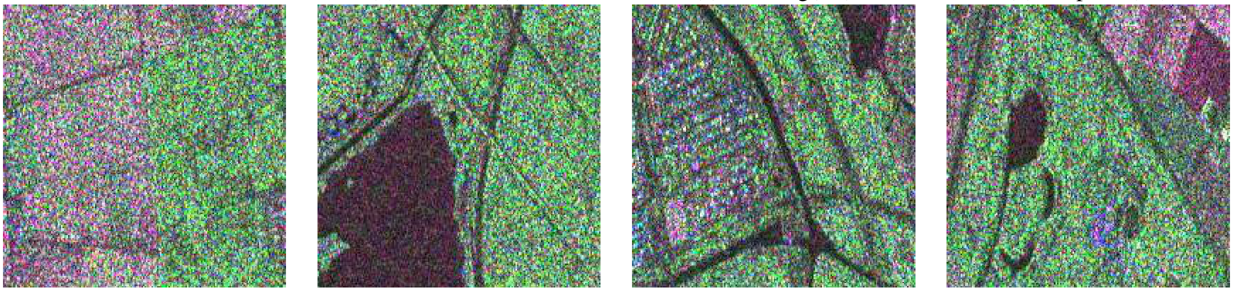
Table 1: Comparison of classification accuracies with different classifers(%)

6 CONCLUSIONS AND FUTURE WORK

This paper presents the utilization of multi-modal markov aspect model for weakly supervised PolSAR image classification.

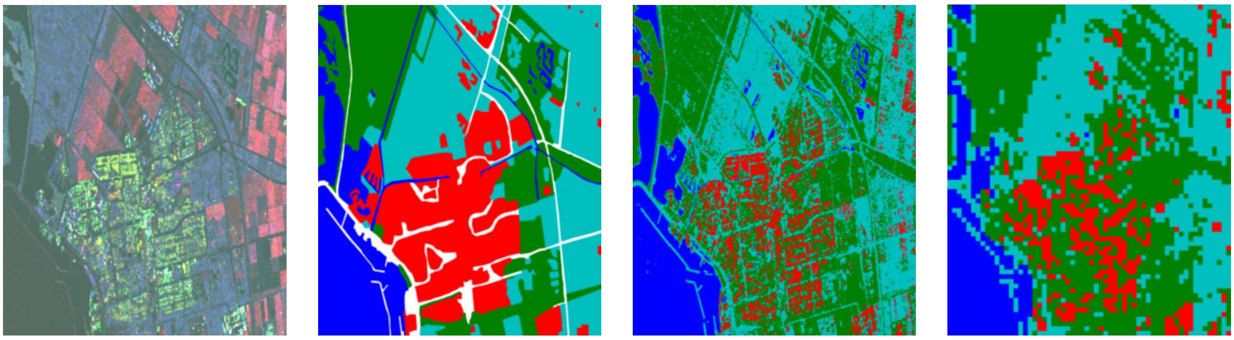


(a) Water(Wa) (b) Woodland(Wo) (c) Building(Bu) (d) Cropland(Cr)

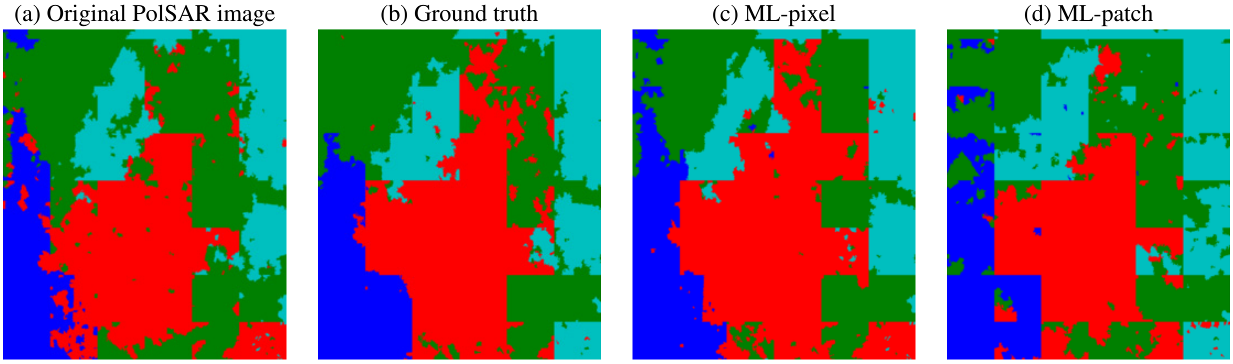


(e) Wo+Cr (f) Wa+Wo (g) Wa+Bu+Wo (h) Cr+Wo+Wa

Figure 2: Some training samples annotated with keywords.



(a) Original PolSAR image (b) Ground truth (c) ML-pixel (d) ML-patch



(e) MMAM-All (f) MMAM-Hist (g) MMAM-GMRF (h) MMAM-Huy

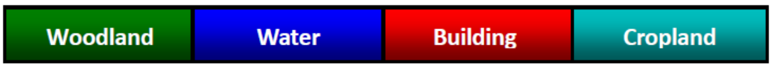
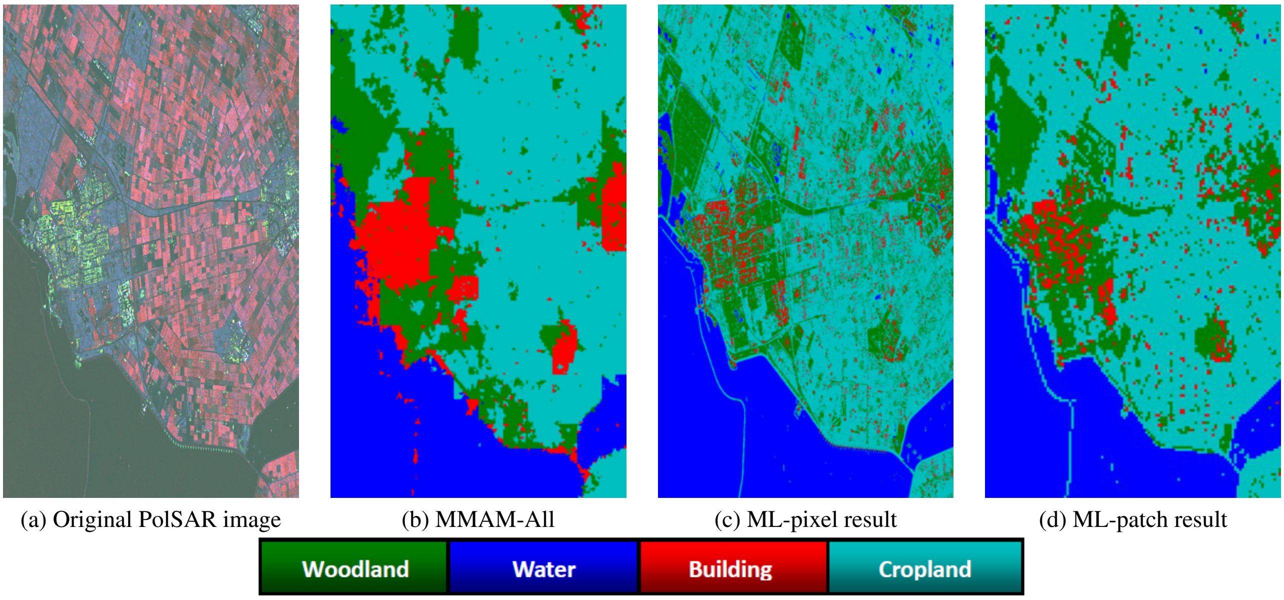


Figure 3: (a) Quantitative evaluation area of 1400 $\times$ 1200 pixels; (b) The corresponding hand-labeled ground truth; (c) Classification result using pixel-based Wishart ML; (d) Classification result using patch-based Wishart ML (patch size:20 $\times$ 20); (e) Classification result using MMAM with all features;(f) Classification result using MMAM with histogram features ;(g) Classification result using MMAM with GMRF features ;(h) Classification result using MMAM with Huynen decomposition features.

It has been tested and validated on a large RadarSat-2 PolSAR scene image classification task, and produces satisfactory classification results, it outperforms traditional Wishart ML methods with detailed pixel-level labeled training data, even when using only one feature-multichannel histogram. Moreover, we use the over-segmentation based soft assignment techniques (Patch to Pixel labels mapping) to reduce the block effect in each subimage and improve the visual effects. While the results presented here are encouraging, there is still a need for further improvements. Future extensions would be the introduction of other sources of contextual information like scale information and the combination with more informative feature descriptors.



(a) Original PolSAR image (b) MMAM-All (c) ML-pixel result (d) ML-patch result

Figure 4: Results of the classification of a large PolSAR image (RadarSat-2 polarimetric SAR data of Flevoland in Netherlands, with size of 4000×2400 pixels) into the four semantic classes: woodland, cropland, water, building.

REFERENCES

Beaulieu, J. M. and Touzi, R., 2004. Segmentation of textured polarimetric sar scenes by likelihood approximation. IEEE Transactions on Geoscience and Remote Sensing, 42(10), pp. 2063–2072.

Blei, D., Ng, A. and Jordan, M., 2003. Latent direchlet allocation. Journal of Machine Learning Research 3(3), pp. 993–1022.

Clausi, D. A., 2001. Comparison and fusion of co-occurrence, gabor, and mrf texture features for classification of sar sea ice imagery. Atmosphere & Oceans, 39(4), pp. 183–194.

Comaniciu, D. and Meer, P., 2002. Mean shift: A robust approach toward feature space analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(5), pp. 603–619.

Ersahin, K., Cumming, I. G. and Ward, R. K., 2010. Segmentation and classification of polarimetric sar data using spectral graph partitioning. IEEE Transactions on Geoscience and Remote Sensing, 48(1), pp. 164–174.

He, W. J., Jager, M., Reigber, A. and Hellwich, O., 2008. Building extraction from polarimetric sar data using mean shift and conditional random fields. The 7th European Conference on Synthetic Aperture Radar (EuSAR), III, pp. 439–442.

Hofmann, T., 2001. Unsupervised learning by probabilistic latent semantic anaylysis. Machine Learning 42(1-2), pp. 177–196.

Huynen, J. R., 1990. The stokes matrix parameters and their interpretation in terms of physical target properties. Journé es Internationales de la Polarimétrie Radar, pp. 524–531.

Lardeux, C., Frison, P.-L., Tison, C., Souyris, J.-C., Stoll, B., Fruneau, B. and Rudant, J.-P., 2009. Support vector machine for multifrequency sar polarimetric data classification. IEEE Transactions on Geoscience and Remote Sensing, 47(12), pp. 4143–4152.

Lee, J. S., Grunes, M. R. and Kwok, R., 1994. Classification of multi-look polarimetric sar imagery-based on complex wishart distribution. International Journal of Remote Sensing, 15(11), pp. 2299–2311.

Lee, J. S., Grunes, M. R. and Pottier, E., 2001. Quantitative comparison of classification capability: Fully polarimetric versus dual and single-polarization sar. IEEE Transactions on Geoscience and Remote Sensing 39(11), pp. 2343–2351.

Li, F.-F. and Perona, P., 2005. A bayesian hierarchical model for learning natural scene categories. Proc. of IEEE Computer Society Conference on. Computer Vision and Pattern Recognition, pp. 524–531.

Liénou, M., Maître, H. and Datcu, M., 2010. Semantic annotation of satellite images using latent dirichlet allocation. IEEE Geoscience and Remote Sensing Letters, 7(1), pp. 28–32.

Shimoni, M., Borghys, D., Heremans, R., Perneel, C. and Acheroy, M., 2009. Fusion of polsar and polinsar data for land cover classification. International Journal of Applied Earth Observation and Geoinformation, 11(3), pp. 169–180.

Verbeek, J. and Triggs, B., 2007. Region classification with markov field aspect models. Proc. of IEEE Computer Society Conference on. Computer Vision and Pattern Recognition,.

Wu, Y. H., Ji, K. F., Yu, W. X. and Su, Y., 2008. Region-based classification of polarimetric sar images using wishart mrf. IEEE Geoscience and Remote Sensing Letters, 5(4), pp. 668–672.

Yedidia, J., Freeman, W. and Weiss, Y., 2005. Constructing free-energy approximations and generalized belief propagation algorithms. IEEE Transactions on Information Theory, 51(7), pp. 2282–2312.

Zou, T. Y., Yang, W., Dai, D. X. and Sun, H., 2010. Polarimetric sar image classification using multi-features combination and extremely randomized clustering forests. EURASIP Journal on Advances in Signal Processing, 2010, pp. 1–9.

ACKNOWLEDGEMENTS

This work was supported in part by the National Natural Science Foundation of China (No.40801183, 60890074), the National High Technology Research and Development Program of China (No.2007AA12Z180, 155) and LIESMARS Special Research Funding.