Bayesian Classification of Hyperspectral Imagery Based on Probabilistic Sparse Representation and Markov Random Field

Linlin Xu, *Student Member, IEEE*, and Jonathan Li, *Senior Member, IEEE*

***Abstract*—This letter presents a Bayesian method for hyperspectral image classification based on the sparse representation (SR) of spectral information and the Markov random field modeling of spatial information. We introduce a probabilistic SR approach to estimate the class conditional distribution, which proved to be a powerful feature extraction technique to be combined with the label prior distribution in a Bayesian framework. The resulting maximum *a priori* problem is estimated by a graph-cut-based α-expansion technique. The capabilities of the proposed method are proven in several benchmark hyperspectral images of both agricultural and urban areas.**

***Index Terms*—Bayes classifier, graph cut, hyperspectral image classification, Markov random field (MRF), probabilistic sparse representation (PSR).**

I. Introduction

THE classification of remotely sensed hyperspectral imagery constitutes a challenging data-mining and machine learning problem due to not only the high dimensionality of various spectral bands but also the ambiguity in spectral signatures of different classes caused by the existence of mixed pixels [1]. In light of these difficulties, one essential issue is how to extract the most compact and discriminative features from the high-dimensional hyperspectral bands. Among many recent studies [1]–[5], the sparse representation (SR) approach has proven to be an extremely powerful tool for hyperspectral image classification [2], [3]. It assumes that the high-dimensional spectral vector can be sparsely represented by a few atoms in a dictionary consisting of training samples. Therefore, forcing sparsity, the training samples in all classes will compete for their involvement in representing the spectral vector. The most relevant class will eventually win large shares, resulting in small representational residual, whereas the wrong or less-relevant classes will have no or less involvement, leading to high representational residual. Therefore, the label of a pixel can be determined by selecting the minimum residuals among all classes. While this approach has proven its capability in revealing the most discriminative information hidden in the high-dimensional spectral vector, there is still a lack of probabilistic approach, which provides the probability features rather than residuals. A probabilistic approach is particularly important, considering the facts that integrating contexture/spatial information is an essential issue for hyperspectral image classification [1]–[3], [5] and employing the Markov random field (MRF) method, i.e., a classic and powerful method for modeling spatial information, requires conditional probability in a Bayesian framework [1], [6], [9], [10].

In this letter, we proposed a probabilistic SR (PSR) approach to be integrated with the MRF technique in a Bayesian framework. Instead of using a unified dictionary consisting training samples from all classes, we design one dictionary for each class. Therefore, we derive a conditional probability for the spectral vector by sparsely representing it over the class-dependent dictionaries. While this probabilistic formulation of SR is used with MRF for hyperspectral data classification, it may also help other statistical methods in other applications. The rest of the letter is organized as follows: Section II discusses the proposed PSR method and its integration with the MRF technique. In Section III, experiments are designed to examine the performance of the proposed method. Section IV concludes this study.

II. Proposed Approach

A. Problem Formulation

In this letter, we denote the discrete lattice spanned by hyperspectral imagery by T and a site in the lattice by $t \in T$. We represent the observation at site t by $\boldsymbol{x}_t$, i.e., a p-dimensional random vector taking on values of various spectral bands, and the label of site t by y_t, i.e., a random variable taking on a class $\{1, \ldots, n\}$. Then, a hyperspectral image can be expressed as $\boldsymbol{x} = \{\boldsymbol{x}_t | t \in T\}$ and the labels of this image as $\boldsymbol{y} = \{y_t | t \in T\}$. In the classification problem, we are trying to infer $\boldsymbol{y}$ based on $\boldsymbol{x}$, which, in the Bayesian framework, can be achieved by maximizing the posterior distribution of $\boldsymbol{y}$ given $\boldsymbol{x}$, i.e.,

$$p(\boldsymbol{y}/\boldsymbol{x}) \propto p(\boldsymbol{x}/\boldsymbol{y})p(\boldsymbol{y}) \tag{1}$$

Manuscript received March 15, 2013; revised July 12, 2013; accepted August 18, 2013.

L. Xu is with the Department of Geography and Environmental Management, University of Waterloo, Waterloo, ON N2L 3G1, Canada (e-mail: l44xu@uwaterloo.ca).

J. Li is with the Key Laboratory of Underwater Acoustic Communication and Marine Information Technology, Ministry of Education, and the School of Information Science and Engineering, Xiamen University, Xiamen 361005, China, and also with the Department of Geography and Environmental Management, University of Waterloo, Waterloo, ON N2L 3G1, Canada (e-mail: junli@xmu.edu.cn; junli@uwaterloo.ca).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LGRS.2013.2279395

where $p(\boldsymbol{x}/\boldsymbol{y})$ denotes the probability distribution of spectral vector $\boldsymbol{x}$ conditioned on $\boldsymbol{y}$, which allows the modeling of spectral information; $p(\boldsymbol{y})$ is the *a priori* probability of labels, which allows the modeling of spatial information.

In this letter, $p(\boldsymbol{x}/\boldsymbol{y})$ is approached by SR to mine the most discriminative information hidden in spectral bands, whereas $p(\boldsymbol{y})$ is implemented by the MRF-based multiple logistic (MLL) prior to constrain regional smoothness. The maximum *a priori* (MAP) problem is solved by the graph-cut-based α-expansion algorithm.

B. PSR

In this letter, we assume that a spectral vector in a class can be sparsely represented by the training samples in the same class. Therefore, as opposed to the classic SR approach that adopts a unified dictionary for all classes [2], [3], we adopt separate dictionaries for different classes. We express the observed signal variable at site t that belongs to class k as

$$\boldsymbol{x}_t^{(k)} = \mathbf{A}^{(k)}\boldsymbol{s}_t^{(k)} + \boldsymbol{n} \tag{2}$$

where $\mathbf{A}^{(k)} = \{\boldsymbol{a}_1^{(k)}, \boldsymbol{a}_2^{(k)}, \ldots, \boldsymbol{a}_{M_k}^{(k)}\}$ is the dictionary consisting of training samples in class k; $\boldsymbol{s}_t^{(k)}$ is the sparse vector corresponding to class k, whose nonzero elements define which columns in $\mathbf{A}^{(k)}$ will be used; and $\boldsymbol{n}$ is the class-independent zero-mean Gaussian noise with diagonal covariance matrix $\boldsymbol{\Lambda}$. Although it is reasonable to assume different $\boldsymbol{n}$ for different classes, it would increase the number of unknown parameters, consequently the risk of overfitting. In our formulation, we assume that $\mathbf{A}^{(k)}\boldsymbol{s}_t^{(k)}$ is capable of capturing the discriminative information in $\boldsymbol{x}_t^{(k)}$; thus, the random noise $\boldsymbol{n}$ is class independent. We treat $\mathbf{A}^{(k)}\boldsymbol{s}_t^{(k)}$ as the fixed effect; hence, the conditional likelihood of spectral vector $\boldsymbol{x}$ can be expressed as

$$\begin{aligned} &p(\boldsymbol{x} = \boldsymbol{x}_t/y = k) \\ &\quad = \frac{1}{(2\pi)^{p/2}|\boldsymbol{\Lambda}|^{1/2}} \\ &\qquad \times \exp\left\{-\frac{1}{2}\left(\boldsymbol{x}_t - \mathbf{A}^{(k)}\boldsymbol{s}_t^{(k)}\right)^T \boldsymbol{\Lambda}^{-1}\left(\boldsymbol{x}_t - \mathbf{A}^{(k)}\boldsymbol{s}_t^{(k)}\right)\right. \end{aligned} \tag{3}$$

$$\boldsymbol{\Lambda} = \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_p \end{bmatrix}. \tag{4}$$

The matrix $\mathbf{A}^{(k)}$ can be implemented as a dictionary storing training samples in class k. Given the dictionary $\mathbf{A}^{(k)}$, the unknown sparse vector $\boldsymbol{s}_t^{(k)}$ can be estimated by solving the following optimization problem:

$$\hat{\boldsymbol{s}}_t^{(k)} = \arg\min \left\|\mathbf{A}^{(k)}\boldsymbol{s}_t^{(k)} - \boldsymbol{x}_t\right\|_2 \quad \text{subject to} \left\|\boldsymbol{s}_t^{(k)}\right\|_0 \leq \tau. \tag{5}$$

The l_0 norm $\|.\|_0$ will simply count the nonzero items in $\boldsymbol{s}_t^{(k)}$. Hence, the optimal $\hat{\boldsymbol{s}}_t^{(k)}$ is estimated by minimizing the representation error with constraint on sparsity level. This NP-hard optimization problem can be solved by some greedy pursuit algorithms, such as orthogonal matching pursuit (OMP) or subspace pursuit. Interested readers are referred to [7] and [8] for further information. The estimation of the second unknown parameter $\boldsymbol{\Lambda}$ relies on the label information. This issue can be solved by the expectation maximization algorithm, by treating the label $\boldsymbol{y}$ as missing information [6]. Therefore, $\boldsymbol{\Lambda}$ is estimated from representation residuals in an iterative manner (see Algorithm 1).

This PSR leads naturally to a discriminative model. Assuming the labels of different sites are independent, according to the Bayes rule, the posterior probability of y_t is given by

$$p(y_t/\boldsymbol{x}_t) \propto p(\boldsymbol{x}_t/y_t)p(y_t). \tag{6}$$

Assuming that the classes are equally likely, then $p(y_t/\boldsymbol{x}_t) \propto p(\boldsymbol{x}_t/y_t)$. Therefore, according to the MAP criterion, we can estimate y_t by maximizing $p(\boldsymbol{x}_t/y_t)$ over different classes. We refer to our classifier as PSR, whose detailed implementation is summarized in Algorithm 1.

Algorithm 1: PSR

Input: training dictionaries for all classes $\{\mathbf{A}^{(1)}, \ldots, \mathbf{A}^{(n)}\}$, data matrix $\boldsymbol{x} = \{\boldsymbol{x}_t | t \in T\}$
Output: class labels $\boldsymbol{y} = \{y_t | t \in T\}$
Initialization: $\hat{\boldsymbol{\Lambda}} = \mathbf{I}$; $i := 1$; $\hat{\boldsymbol{s}}_t^{(k)} = OMP(\mathbf{A}^k, \boldsymbol{x}_t, \tau)$; for $k = 1, 2, \ldots, n$ and $t \in T$
while $i \leq iters$ or $sum(diag|\hat{\boldsymbol{\Lambda}}^i - \hat{\boldsymbol{\Lambda}}^{i-1}) > s$ **do**
  $\hat{y}_t = \arg\min_{y_t}\{-\log(p(\boldsymbol{x}_t/y_t))\}$
  $\hat{\boldsymbol{\Lambda}}^i = \text{var}(\{\boldsymbol{x}_t - \mathbf{A}^{\hat{y}_t}\boldsymbol{s}_t^{\hat{y}_t} | t \in test\ set\})$
end while

C. MRF-Based MLL Prior

Although PSR itself constitutes a classifier, it ignores the contextual information that is of great importance for hyperspectral data classification. Therefore, we further incorporate the spatial information by using the MRF-based MLL prior. The MRF is a classical method for modeling contextual information [9]. It promotes identical class labels for spatially close pixels. The MRF-based approach is often implemented by the MLL model, which can be expressed as [10]

$$p(\boldsymbol{y}) = \frac{1}{z}\exp\left(-\sum_{t\in T}\sum_{u\in N_t}\delta(y_t, y_u)\right) \tag{7}$$

where N_x denotes the neighborhood centered at site t; and $\delta(y_t, y_u) = -1$ if $y_t = y_u$, whereas $\delta(y_t, y_u) = 1$ if $y_t \neq y_u$.

D. Complete Algorithm

The PSR and MLL in Section II-B and C are incorporated into a Bayesian framework and solved by the MAP criterion.

TABLE I
OA, AA, AND κ STATISTICS OBTAINED BY DIFFERENT METHODS (BEST RESULTS ARE HIGHLIGHTED IN BOLD TYPEFACE)

Classifies	Datasets								
	Indian Pines			Pavia U			Pavia C		
	OA(%)	AA(%)	κ	OA(%)	AA(%)	κ	OA(%)	AA(%)	κ
OMP	67.8	64.7	0.632	80.4	83.1	0.738	96.2	91.1	0.931
OMPMLL	67.9	64.7	0.632	80.4	83.1	0.738	96.3	91.5	0.932
MLRsub	70.5	68.5	0.663	76.2	77.9	0.701	94.6	84.7	0.897
MLRsubMLL	94.7	**90.6**	0.946	96.1	95.2	0.953	98.3	95.8	0.970
PSR1	67.0	56.8	0.623	77.9	78.4	0.703	93.9	83.6	0.889
PSR1MLL	93.7	75.6	0.928	98.4	98.2	0.979	**99.5**	**98.2**	**0.990**
PSR2	72.3	65.1	0.686	78.4	78.3	0.709	93.7	82.6	0.884
PSR2MLL	**97.8**	83.5	**0.975**	**99.1**	**98.8**	**0.987**	99.4	97.9	0.989

The optimal labeling $\hat{y}$ can be obtained according to MAP criterion

$$\hat{\boldsymbol{y}} = \arg\min_{\boldsymbol{y}} \left\{ \sum_{t\in T} \left[-\log p\left(\boldsymbol{x}_t | y_t, \Lambda, \boldsymbol{s}_t^{(y_t)}\right) + \gamma \sum_{u\in N_t} \delta(y_t, y_u) \right] \right\} \tag{8}$$

where γ is the weighting parameter that determines the relative contribution of the two components. This combinational optimization problem of estimating $\boldsymbol{y}$ given $\boldsymbol{\Lambda}$ and $\boldsymbol{s}_t^{(y_t)}$ is solved in this letter by the graph-cut-based α-expansion algorithm, which proved being capable of providing efficient and effective approximation to the MAP segmentation in computer vision [11], [12]. We refer to the complete algorithm in this section as PSRMLL, whose detailed implementation is summarized in Algorithm 2. The time complexity of PSRMLL is largely determined by the complexity of the OMP algorithm, i.e., $O(\tau p M)$ with M being the number of atoms in dictionary, and the complexity of the α-expansion algorithm, i.e., $O(T)$ with T being the number of pixels.

Algorithm 2: PSRMLL

Input: training dictionaries for all classes $\{\mathbf{A}^{(1)}, \ldots, \mathbf{A}^{(n)}\}$, data matrix $\boldsymbol{x} = \{\boldsymbol{x}_t | t \in T\}$
Output: class labels $\boldsymbol{y} = \{y_t | t \in T\}$
Initialization: $\boldsymbol{\Lambda} = \mathbf{I}$; $i := 1$; $\hat{\boldsymbol{s}}_t^{(k)} = OMP(\mathbf{A}^{(k)}, \boldsymbol{x}_t, \tau)$ for $k = 1, 2, \ldots, n$ and $t \in T$
while $i \leq iters$ or $sum(diag|\hat{\boldsymbol{\Lambda}}^{i} - \hat{\boldsymbol{\Lambda}}^{i-1}) > s$ **do**
$\mathbf{P} = \{p(\boldsymbol{x}_t / y_t = k) | k = 1, 2, \ldots, n \text{ and } t \in T\}$
$\hat{\boldsymbol{y}} = \alpha - \text{expansion } (\mathbf{P}, \gamma)$
$\hat{\boldsymbol{\Lambda}}^{i} = \text{var}(\{\boldsymbol{x}_t - \mathbf{A}^{(\hat{y}_t)} \boldsymbol{s}_t^{(\hat{y}_t)} | t \in test\ set\})$
end while

III. EXPERIMENTS

We adopt three benchmark hyperspectral images, i.e., Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) Indian Pines, University of Pavia, and the Center of Pavia (refer to [13] for detailed information), to test the proposed algorithms. The first image was captured by AVIRIS over a vegetation area in Northwestern Indiana, USA, with a spatial resolution of 20 m, consisting of 145 $\times$ 145 pixels of 16 classes and 200 spectral reflectance bands after removing 20 water absorption bands (104–108, 150–163, and 220). The other two hyperspectral images are urban images acquired by the Reflective Optics System Imaging Spectrometer, with a spatial resolution of 1.3 m, consisting of 103 spectral bands after removing 12 noisy bands. The University of Pavia image is centered at the University of Pavia, consisting of 610 $\times$ 340 pixels, whereas the Center of Pavia image is at the center of Pavia City, consisting of 1096 $\times$ 492 pixels. Both images have nine ground-truth classes.

A. Design of Experiments

We implemented Algorithms 1 and 2 in Section II-B and D, which are referred to as PSR2 and PSR2MLL. To examine the influence of $\boldsymbol{\Lambda}$, we forced $\hat{\boldsymbol{\Lambda}}$ in PSR2 and PSR2MLL to be the unit matrix. Therefore, the resulting algorithms are referred to as PSR1 and PSR1MLL, respectively. We experimentally set $iter = 20$ and $s = 0.1$ for PSR2 and PSR2MLL and $\gamma = 20$, $\tau = 5$ for all proposed algorithms. In Section III-D, we will explore the sensitivity of these parameters. We also implemented the OMP algorithm in [2] and adopted the residuals in OMP as the data cost to feed the α-expansion algorithm (referred to as OMPMLL). Moreover, since the MLRsubMLL approach in [1] is also an MRF-based approach, we included this algorithm along with the MLRsub for a comparison study. The smooth cost in MLRsubMLL was set to be 2 for optimal performance, whereas all other parameters followed [1].

For the labeled pixels in these data sets, we randomly select a certain number of pixels from each class as training samples, whereas the other labeled pixels are used as the test set. For the Indian Pines data set, training samples in each class constitute 10% of the total samples in that class. For the other two data sets, we adopt a popular approach, and the number of training samples in each class is the same as that in [2]. For further details, the reader is referred to [2].

To be consistent with the other researchers, we adopt three numerical measures, i.e., overall accuracy (OA), average accuracy (AA), and the κ coefficient, for evaluation purposes [12]. To account for the possible bias produced by random sampling, each experiment is performed ten times on different sampling results. The numerical values in Table I are the average of the ten realizations. However, the maps in Fig. 1 are from one realization.

B. Numerical Comparison

Table I provides the statistics of different algorithms on three benchmark data sets. Overall, PSR2MLL greatly outperformed

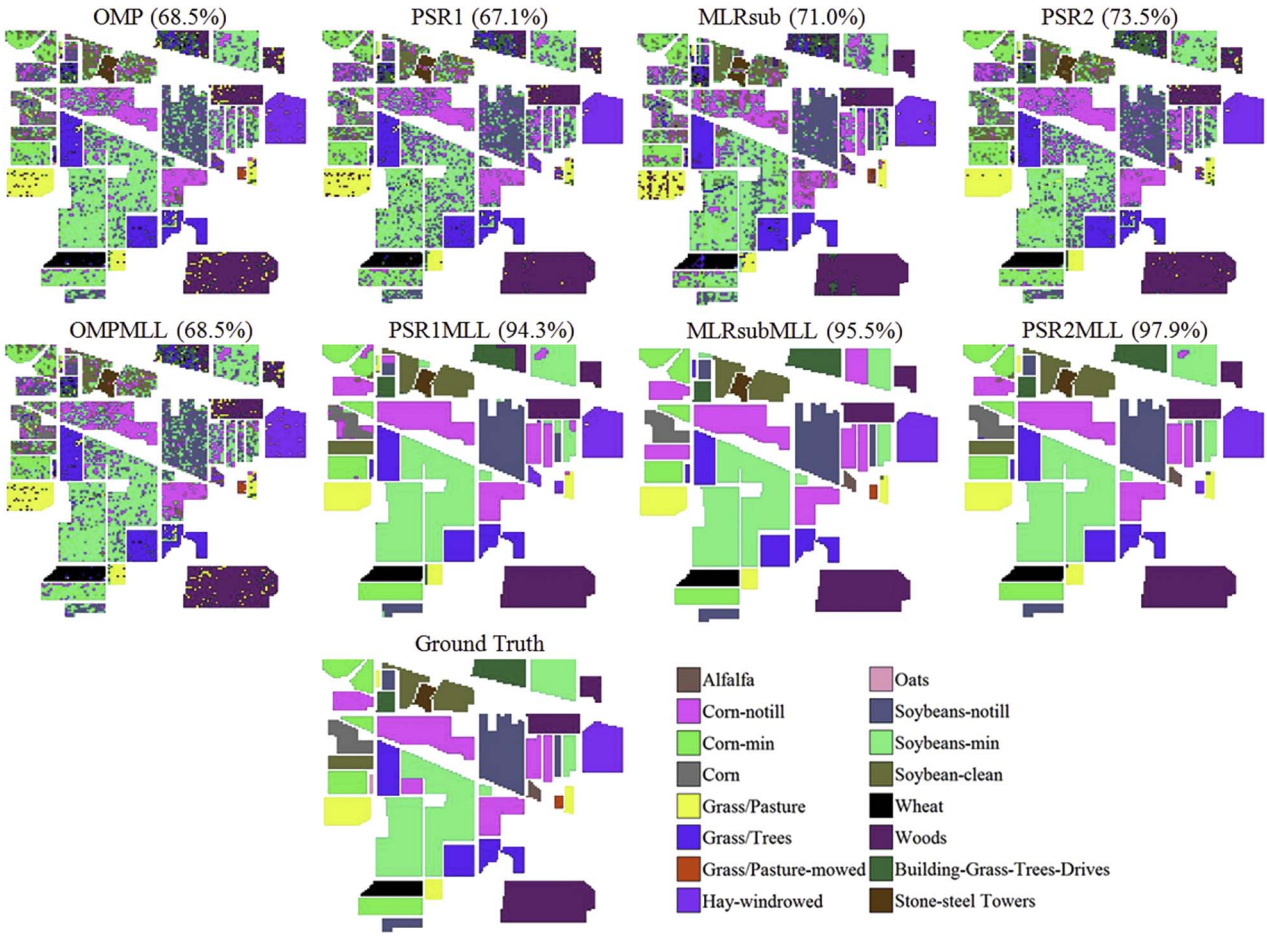

Fig. 1. Classification maps obtained by different methods on the AVIRIS Indian Pines data set (OA is reported in the parentheses).

the other approaches on most data sets, achieving OA of 97.8%, 99.1%, and 99.4%, respectively.

Comparing with PSR1 and PSR2, the OAs of PSR1MLL and PSR2MLL increased on average 25%, 21%, and 6% on the three data sets, respectively, indicating the importance and benefit of integrating the SR-based classifier with MRF to utilize both spectral and spatial information for hyperspectral image classification. MLRsubMLL also significantly increased the performance of MLRsub. However, nearly no performance increase for OMPMLL over OMP was observed. It is mostly because OMP is a hard classifier, which produces residual features rather than probability features.

Comparing with PSR1 and PSR1MLL, PSR2 and PSR2MLL achieved higher OA on Indian Pines, slightly higher values on Pavia U, and comparable values on Pavia C. These results justify the idea of accounting for the variance heterogeneity across different spectral bands. Moreover, they may also indicate that addressing variance inhomogeneity is more beneficial when the quality of training samples is low, considering that the Indian Pines data set, on which the PSR2 and PSR2MLL achieve a higher performance increase than on the other two data sets, assumes higher dimensionality due to more spectral bands, heavier mixed pixel effect caused by lower spatial resolution, and smaller number of training samples in most classes than Pavia U and particularly Pavia C.

It is desirable to compare PSR2MLL and MLRsubMLL, since both approaches are MRF-based generative models for MAP classification. PSR2 slightly outperformed MLRsub on Indian Pines and Pavia U, whereas MLRsub achieves better results on Pavia C. Nevertheless, the adoption of MLL prior enabled PSR2MLL to achieve higher OA and κ values on all data sets.

C. Visual Comparison

Fig. 1 shows the classification maps by different algorithms on the Indian Pines image. Generally speaking, it indicates consistent results with the numerical measures. As we can see, algorithms without MLL prior, i.e., OMP, PSR1, PSR2, and MLRsub, produced intense artifacts in the classification map due to the existence of mixed pixels in the image. Although all four algorithms performed seemingly well, careful inspection indicates that PSR2 yields fewer artifacts than the others in certain classes, e.g., Grass/Pasture, Building-Grass-Tree-Drives, and Soybeans-min. By combining with MLL prior, PSR1MLL, PSR2MLL, and MLRsubMLL produced very smooth results, although there still exist misclassified patches in classes such as Soybeans-min and Building-Grass-Tree-Drives. Nevertheless, some small classes, such as Oats, were totally misclassified because of the lack of enough training

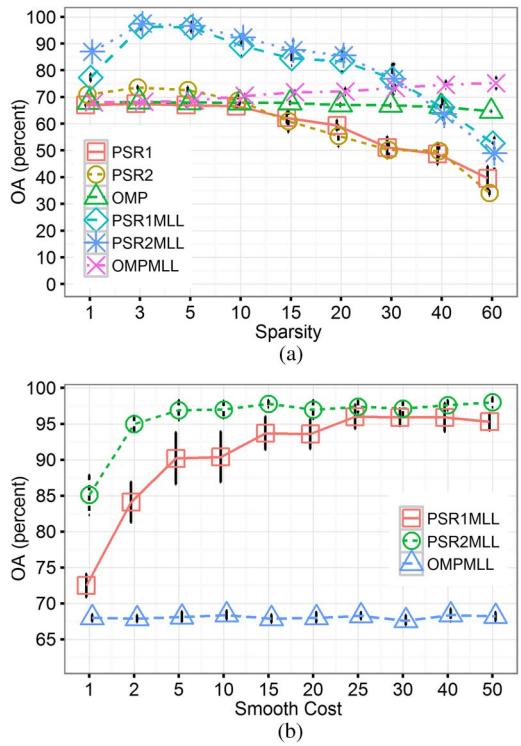

Fig. 2. Error bar of OA as a function of (a) sparsity τ and (b) smooth cost γ.

samples for small classes. We also noticed that there is not much difference between the maps of OMPMLL and OMP.

D. Sensitivity of Parameters

This section explored the sensitivity of two important parameters, i.e., sparsity level and smooth cost for SR-based algorithms. Fig. 2 plots the error bar of OA as a function of sparsity τ and smooth cost γ based on the AVIRIS Indian Pines data set.

Fig. 2(a) indicates that PSR-based algorithms achieved the highest performance when the sparse level was 3. Therefore, from the sparsity level of 3, the performance of PSR-based algorithms reduced quite sharply. This is not surprising because the increased sparsity level allows the wrong class to represent the test sample equally well as the true class, which consequently leads to the loss of discriminative power. PSR2MLL achieved higher OA than PSR1MLL, and both PSR1MLL and PSR2MLL outperform OMPMLL when the sparsity level is lower than 30. OMP achieved stable results, and OMPMLL demonstrated slightly increased performance on high sparsity level.

In Fig. 2(b), the increase in smooth cost increased the performance of PSR1MLL and PSR2MLL to a stable level but did not indicate the noticeable influence on OMPMLL. Moreover, PSR2MLL achieved higher accuracy but lower variance than PSR1MLL across most smooth-cost levels, indicating the worth of accounting for the variance heterogeneity in PSR.

IV. Conclusion

In this letter, we have proposed a PSR approach to be integrated with MRF in a Bayesian framework for hyperspectral image classification. We assume that the spectral vector in a class can be sparsely represented by the training samples in the same class. Moreover, the representation error is assumed being class independent, with zero mean and diagonal covariance matrix. Based on these assumptions, we have derived the class conditional distribution of the spectral vector, which is used with the MRF label prior distribution to form a MAP problem. The proposed approach is solved by graph-cut-based α-expansion techniques. On benchmark hyperspectral images, the proposed algorithm achieved new state-of-the-art performance.

References

[1] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Spectral-spatial hyperspectral image segmentation using subspace multinomial logistic regression and Markov random field," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 3, pp. 809–823, Mar. 2012.

[2] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Hyperspectral image classification using dictionary-based sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 10, pp. 3973–3985, Oct. 2011.

[3] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Hyperspectral image classification via kernel sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 217–231, Jan. 2013.

[4] J. Xia, P. Du, X. He, and J. Chanussot, "Hyperspectral remote sensing image classification based on rotation forest," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 1, Jan. 2014, to be published. [Online]. Available: http://ieeexplore.ieee.org

[5] G. Camps-Valls, N. Shervashidze, and K. M. Borgwardt, "Spatio-spectral remote sensing image classification with graph kernels," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 4, pp. 741–745, Oct. 2010.

[6] H. Deng and D. A. Clausi, "Unsupervised segmentation of synthetic aperture radar sea ice imagery using a novel Markov random field model," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 528–538, Mar. 2005.

[7] J. Tropp and A. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.

[8] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *IEEE Trans. Inf. Theory*, vol. 55, no. 5, pp. 2230–2249, May 2009.

[9] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-6, no. 6, pp. 721–741, Nov. 1984.

[10] S. Li, *Markov Random Field Modeling in Image Analysis*. New York, NY, USA: Springer-Verlag, 2001.

[11] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 11, pp. 1222–1239, Nov. 2001.

[12] S. Bagon, Matlab wrapper for graph cut, Dec. 2006. [Online]. Available: http://www.wisdom.weizmann.ac.il/~bagon

[13] *Hyperspectral Remote Sensing Scenes*, Mar. 2013. [Online]. Available: http://www.ehu.es/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes