2014 22nd International Conference on Pattern Recognition

Local Label Probability Propagation for Hyperspectral Image Classification

Haichang Li, Jiangyong Duan, Shiming Xiang, Lingfeng Wang and Chunhong Pan
Institute of Automation, Chinese Academy of Sciences
{hcli, jyduan, smxiang, lfwang, chpan}@nlpr.ia.ac.cn

***Abstract*—Classification of hyperspectral images is an important issue in remote sensing image processing systems. Hyperspectral images have advantages in pixel-wise classification owing to the high spectral resolution. However, the pixel-wise classification result often introduces the salt-and-pepper appearance because of the complex noise produced by atmosphere and instrument. An effective way to overcome this phenomenon is to resort to the spatial information. This paper proposes a method to solve the above problem by using spatial similarity information. First, in order to avoid the effect of noisy pixels and mixed pixels, reliable seeds are selected in local windows according to the agreement between the central pixel and its spatial neighbors. Then, the information of the reliable seeds is propagated to their spatial neighbors by a graph Laplacian. Specifically, the graph Laplacian is designed to propagate information among spatial neighbors with close similarity relationship so that some small or long thin objects are identified. Through the seed selection and local reliable information propagation, the problem of noisy labels is solved elegantly. Experiments on three real hyperspectral data sets with different spatial resolution, spectral resolution and land covers demonstrate the effectiveness of our method.**

I. Introduction

Nowadays, we can acquire images in very narrow spectrum such as 10 nanometers by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS). Such high spectral resolution makes absorption features of different land covers expressed, which helps to identify objects. Thus, hyperspectral image classification has been a hot topic for several years.

Generally speaking, the hyperspectral image classification can be divided into two groups: pixel-wise classification and spatial-spectral classification. For the pixel-wise classification, each pixel with its wavebands recorded in hyperspectral images is considered as a sample. Thus, the pixel-wise hyperspectral image classification is addressed as a classical pattern classification problem. Many traditional classification methods have been introduced into this field, including the Maximum Likelihood Classification (MLC) algorithm [1], the *K*-Nearest Neighbor (*K*NN) classification algorithm [2] and the kernel based methods [3]. Among these methods, kernelized Support Vector Machine (SVM) has been proved to be very effective in pixel-wise hyperspectral image classification, representing the state-of-the-art approaches. It transforms the original data into a kernel feature space, then finds the decision hyperplane by maximizing the margin between different classes. The main shortcomging of the pixel-wise classification is that the spatial information is not utilized to help improve the performance of classification.

Recently, the spatial information has been introduced into hyperspectral image classification. One way to incorporate the spatial information is in the feature level. Benediktsson et al. [4] propose the Extended Morphological Profiles (EMP) method to integrate the spectral information and the spatial information together. They use the morphological transformation to construct a Morphological Profile (MP) on each of the significant principal components for hyperspectral images. Then, all the profiles are concatenated into one extended MP, which is classified by a neural network. Mathieu Fauvel et al. [5] fuse the EMP features and the original spectral features together to build more informative features. Another typical way to incorporate the spatial information is voting in homogenous regions obtained by over-segmentations [6], [7], [8], [9], which achieves high classification accuracy. However, over-segmentation is still an un-solved problem.

Another family to combine the spectral information and the spatial information is Markov Random Field (MRF) [10], [11]. Zhang et al. [10] provide a method to manage the trade-off between the spectral contribution and the spatial contribution in a MRF. Jia et al. [11] introduce a relative homogeneity index to determine the suitable weighting coefficient of the spatial contribution. Besides, other graph-based methods also have been proposed to use the contextual information. For example, Camps-Valls et al. [12] incorporate the contextual information by a graph on the original labeled and unlabeled data with $\{1, 0\}$ form. Bai et al. [13] employ graph cut theory to handle the classification of hyperspectral images. They first implement the pixel-wise classification on the whole image by SVM. Then, pixels with high label probabilities are chosen as seeds in a MRF. Finally, the hyperspectral image is segmented by graph cut algorithm. These graph based methods utilize the contextual information on the $\{1, 0\}$ labeled data. An alternative way is to handle the pixel-wise probability classification result by a graph, which is more accurate and computationally efficient.

This paper presents a graph-based method based on the probability estimates by SVM for hyperspectral image classification. Because of the high spectral resolution, the pixel-wise classification of hyperspectral images is relatively reliable except some noisy labels. Based on this point, we originally employ graph Laplacian on the label probability map estimated by SVM in hyperspectral image classification. First, each pixel is assigned a probability for every class by SVM. The pixels that have labels consistent with the labels of their most spatial neighbors are considered as reliable seeds. Then the probabilities of the reliable seeds propagate to their locally spatial neighbors by a graph Laplacian. Through reliable label probability propagation, the spectral information and the spatial information are used at the same time. Finally, our model is

1051-4651/14 $31.00 © 2014 IEEE
DOI 10.1109/ICPR.2014.729

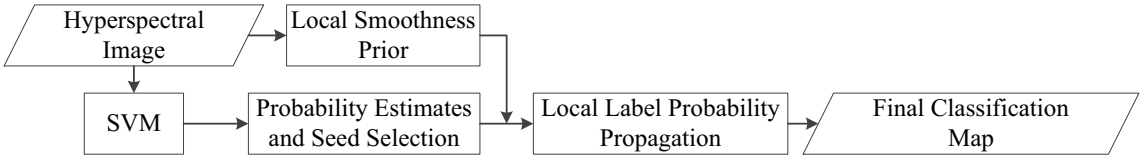


Fig. 1: Flow chart of our algorithm.

optimized with a closed-form solution elegantly and efficiently by solving a sparse linear equation system. Experiments on three real hyperspectral data sets evaluate the effectiveness and efficiency of our method. Specifically, the main contributions of this paper are listed as follows:

- Graph Laplacian is constructed on the label probability map estimated by SVM in hyperspectral image classification. The graph is designed locally with local scaling parameters so that the similarity relationships among local pixels are distinct;
- The pixels with estimated labels that are consistent with most of their spatial neighbors' are selected as reliable seeds. The label information of these reliable seeds is propagated to their neighbors with close similarity relationship so that the propagation is more reasonable and the noisy information is suppressed.

The remainder of this paper is outlined as follows. Section II describes the model. The incorporation of spatial information and optimization process for our model is also described in this section. Experimental results are reported in Section III. Finally, conclusions and future work are drawn in Section IV.

II. Model

Given a scene of hyperspectral images consisting of N pixel vectors, our task is to assign a class label within the c classes of interest for each pixel vector. It is considered as a reliable label probability propagation problem.

The whole flow chart of our scheme is shown in Fig. 1. First, we use SVM [14], [15] to give the initial label probability estimate for each pixel vector in hyperspectral images, i.e., the probability estimate of each sample belonging to the assigned classes. Then, reliable seeds are chosen in local regions, and their label probabilities are propagated to the spatial neighbors by a local smoothness prior. The details are listed as follows.

A. Local Smoothness Prior

Usually, the land covers are continuously distributed in real world. A natural assumption is that the spatial neighbors those are similar with each other in hyperspectral images should be of the same objects.

Suppose two spatial neighborhood pixel vectors are $\mathbf{x}_i \in \mathbb{R}^d$ and $\mathbf{x}_j \in \mathbb{R}^d$, their corresponding label probabilities are $\mathbf{y}_i \in \mathbb{R}^c$ and $\mathbf{y}_j \in \mathbb{R}^c$ respectively, where the k-th element of $\mathbf{y}_i$ corresponds to the probability that $\mathbf{x}_i$ belongs to the k-th class. If $\mathbf{x}_i$ and $\mathbf{x}_j$ is near in the $\mathbb{R}^d$ space, then $\mathbf{y}_i$ and $\mathbf{y}_j$ should be close in the $\mathbb{R}^c$ space. That is $\|\mathbf{y}_i - \mathbf{y}_j\|_2^2$ should be small. In machine learning research, this character is often called as *local smoothness assumption* [16]. This character makes that the rightly classified pixel vectors can provide beneficial information for the miss-classified spatial neighbors. Therefore, we can employ it to obtain a more comfortable thematic map when a relatively good pixel-wise classification map has been obtained.

Denoting the similarity between $\mathbf{x}_i$ and $\mathbf{x}_j$ as w_{ij}, which is defined as follows:

$$w_{ij} = \begin{cases} e^{-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|_2^2}{\sigma}}, & \text{if } j \in \mathcal{N}(i), \\ 0, & \text{otherwise,} \end{cases}$$

where $\mathcal{N}(i)$ denotes the i-th pixel's spatial neighborhood set that have relatively higher similarity with $\mathbf{x}_i$. σ is the heat kernel parameter that is related to the local variation information in this paper. These two characters make label probability just propagate between the more similar pixel vectors in local regions, which can retain long thin objects such as roads in hyperspectral image. In order to make $w_{ij} = w_{ji}$, we substitute their mean for w_{ij} and w_{ji}.

Then, the local smoothness assumption is described as the following constraint,

$$\sum_{i=1}^{N} \sum_{j \in \mathcal{N}(i)} w_{ij} \|\mathbf{y}_i - \mathbf{y}_j\|_2^2 \quad = \quad \sum_{i,j} w_{ij} \|\mathbf{y}_i - \mathbf{y}_j\|_2^2. \quad (1)$$

In terms of matrices, (1) can be reformulated as

$$\begin{aligned} \frac{1}{2} \sum_{i,j} w_{ij} \|\mathbf{y}_i - \mathbf{y}_j\|_2^2 &= \mathrm{Tr}(\mathbf{Y}(\mathbf{D} - \mathbf{W})\mathbf{Y}^T) \\ &= \mathrm{Tr}(\mathbf{Y}\mathbf{L}\mathbf{Y}^T) \end{aligned} \quad (2)$$

where $\mathrm{Tr}(\cdot)$ denotes the trace of a matrix and $\mathbf{W} \in \mathbb{R}^{N \times N}$ is a sparse symmetric matrix with its element w_{ij} representing the similarity between the pixel vectors $\mathbf{x}_i$ and $\mathbf{x}_j$. $\mathbf{D} \in \mathbb{R}^{N \times N}$ is a diagonal matrix with the i-th diagonal element as $D_{ii} = \sum_{j=1}^{N} w_{ij}$. $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_N] \in \mathbb{R}^{c \times N}$ is the final label probability matrix that needs to be computed. $\mathbf{L}$ is a symmetric graph Laplacian matrix [17], which combines the local smoothness and global information together.

B. Selection of Reliable Classified Pixels

The pixel-wise classification result of hyperspectral images determined by SVM is not always reliable. Firstly, the hyperspectral image has complex noise produced by atmosphere and instrument, which makes some clear pixel vectors obscure. Secondly, mini-objects less than a pixel in hyperspectral image and the border pixels between different land covers should produce mixing elements, which makes it hard to determine their exact labels. Since these pixel vectors are unreliable, a

reasonable substitution is that their labels should be the same as the labels of their similar spatial neighbors according to the local smoothness assumption.

As in II-A, we assume that the pixel values of the hyperspectral image are continuous in local regions. Thus, their labels should be locally smooth. The majority votes of labels in a local region are considered reliable. We choose a 3×3 window to determine the reliability of the central pixel vector. If the central pixel's label is consistent with most of its eight spatial neighbors', then it is considered as the reliable pixel, otherwise, it is seen as the unreliable pixel.

Let $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N] \in \mathbb{R}^{d \times N}$ collect all the N pixel vectors in hyperspectral images, and $\mathbf{P} = [\mathbf{p}_1, \mathbf{p}_2, ..., \mathbf{p}_N] \in \mathbb{R}^{c \times N}$ are the corresponding probabilities estimated by SVM, where c is the number of classes, $\mathbf{p}_i$ is the image vector of $\mathbf{x}_i$, and the k-th element of $\mathbf{p}_i$ corresponds to the probability that $\mathbf{x}_i$ belongs to the k-th class. Denote the final classification result as $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_N] \in \mathbb{R}^{c \times N}$, and an indicator vector $\mathbf{s} = [s_1, s_2, ..., s_N]^T \in \mathbb{R}^N$ consists of 1 and 0, representing whether the pixels are reliable or not respectively. Then, the probability reconstruction error for all the reliable pixels is:

$$\sum_{i=1}^{N} s_i \|\mathbf{p}_i - \mathbf{y}_i\|_2^2 \quad = \quad \mathrm{Tr}[(\mathbf{P} - \mathbf{Y})\mathbf{S}(\mathbf{P} - \mathbf{Y})^T] \qquad (3)$$

where $\mathbf{S} \in \mathbb{R}^{N \times N}$ is a diagonal matrix and its diagonal vector is $\mathbf{s}$. The above equation represents the difference between the probabilities estimated by SVM and the final computed probabilities for all the reliable pixel vectors.

C. Local Label Probability Propagation

With the pixel-wise classification result $\mathbf{P} \in \mathbb{R}^{c \times N}$,the seed selection matrix $\mathbf{S}$ and the local smoothness prior, our model is

$$\min_{\mathbf{Y}} \mathrm{Tr}[(\mathbf{P} - \mathbf{Y})\mathbf{S}(\mathbf{P} - \mathbf{Y})^T] + \lambda \mathrm{Tr}(\mathbf{Y}\mathbf{L}\mathbf{Y}^T) \qquad (4)$$

where λ is a trade-off parameter, which controls the reliability between the probability estimates by SVM and the local smoothness prior of labels. This also balances the weights between spectral information and spatial information. The first term makes the final classification map is close to the pixel-wise classification map by SVM, which is determined by the spectral information. The second term locally smoothes the whole classification map, which depends on the spatial information.

Denoting Eqn. (4) as $F(\mathbf{Y})$, taking derivative of $F(\mathbf{Y})$ with respect with $\mathbf{Y}$, then we obtain

$$\frac{\partial F(\mathbf{Y})}{\partial \mathbf{Y}} \quad = \quad 2(\mathbf{Y} - \mathbf{P})\mathbf{S} + 2\lambda \mathbf{Y}\mathbf{L}.$$

Let $\frac{\partial F(\mathbf{Y})}{\partial \mathbf{Y}} = 0$, it follows

$$\mathbf{Y} \quad = \quad \mathbf{P}\mathbf{S}(\mathbf{S} + \lambda \mathbf{L})^{-1}. \qquad (5)$$

The matrix $\mathbf{S} + \lambda \mathbf{L}$ is sparse, positive semi-definite and symmetric, thus the global optimum value of Eqn. (4) can be found efficiently by solving a sparse linear equation system. Since we choose seeds in local regions and the probabilities of the seeds propagate to their similar spatial neighbors, we call this process as Local Label Probability Propagation (LLPP). The index corresponding to the largest entry in each column of $\mathbf{Y}$ is considered as the final label of this column of sample.

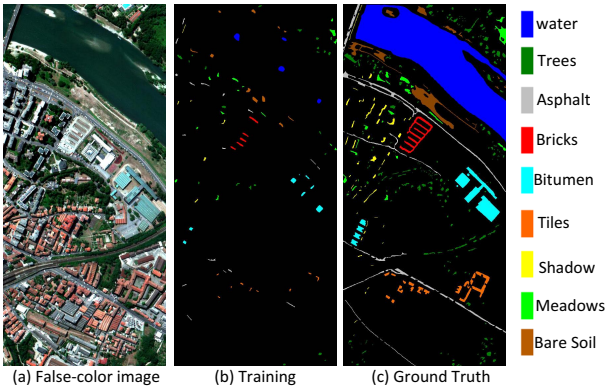


Fig. 2: Center of Pavia image. (a) A three-band false-color image. (b) Training set. (c) Ground Truth.

Through the LLPP, we transform the pixel-wise classification map of SVM to a more smooth and more accurate thematic map. This process removes noise in the classification map through locally reliable label probability propagation using the spatial similarity property. From the perspective of Random Walk [18], for a pixel vector $\mathbf{x}_i$ with probability $\mathbf{p}_i$, the one-step transition probability from $\mathbf{p}_i$ to its spatial neighbor's probability $\mathbf{p}_j$ is proportional to w_{ij} and is given by $p_{ij} = \frac{w_{ij}}{D_{ii}}$. That is, the label probability propagates from a seed to its similar spatial neighbors, and the one-step transition probability is proportional to the similarity between them.

III. Experiments

To evaluate the effectiveness of our method, we compare the following three methods:

- **SVM**[1]: Pixel-wise classification by SVM with Gaussian Radial Basis Function (RBF) kernel [3], [15], which represents the state-of-the-art approaches for pixel-wise hyperspectral image classification.
- **SVM/MV**: Majority Voting within local regions on the classification map of SVM. The label of each pixel is determined by the majority labels in its 3×3 window in our experiments.
- **SVM/LLPP**: Local Label Probability Propagation, our method.

Three data sets acquired by two type of sensors are used in our experiments. The data sets and classification results are described in the following subsections.

A. Data Description

Center of Pavia image: the Center of Pavia image is captured by the Reflective Optics System Imaging Spectrometer (ROISIS-03). A part of the image is used in our experiments, which has 1096×492 pixels with a spatial resolution of $1.3m$

[1]Available at http://www.csie.ntu.edu.tw/ cjlin/libsvm.

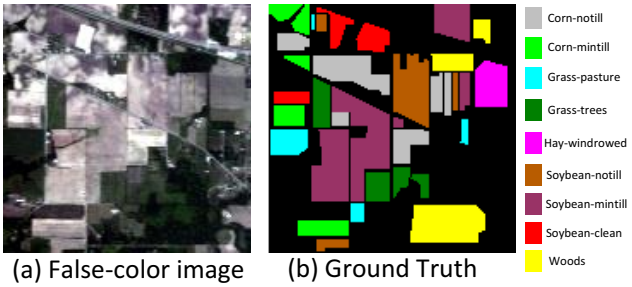


Fig. 3: Indian Pines data. (a) A three-band false-color image. (b) Ground Truth.

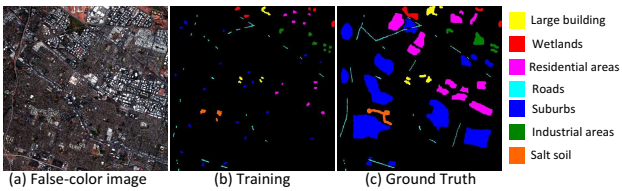


Fig. 4: Moffett Field data. (a) A three-band false-color image. (b) Training set. (c) Ground Truth.

per pixel. The spectral range in the original image is from $0.43\mu m$ to $0.86\mu m$. After removing thirteen water absorption and too noisy spectral bands, only 102 wave-bands are left. A three-band false color image and the reference data are shown in Fig. 2, where nine classes of interest are labeled: water, trees, asphalt, bricks, bitumen, tiles, shadow, meadows and bare soil. The number of training and testing samples are shown in Table. I.

Indian Pines image: the Indian Pines image is recorded by the Airborne Visible/Infrared Imaging Spectrometer(AVIRIS). It has 145×145 pixels with a spatial resolution of about 20m per pixel. There are 220 wave-bands in the original recorded image. After discarding water absorption and too noisy wave-bands, only 200 wave-bands are used in the experiments. A three-band false color image and the reference data are shown in Fig. 3, where only nine classes with more samples are labeled: corn-no till, corn-min till, hay-windrowed, soybeans-no till, soybean-min till, soybean-clean till, grass/pasture, grass/trees and woods. 25% of the samples for each class from the reference data are randomly chosen as the training samples. The remaining samples in the reference data compose the testing set. The detail information is shown in Table. II.

Moffett Field Data[2]**:** the Moffett Field data is recorded by the AVIRIS sensor. The image has 500 $\times$ 500 pixels. After discarding water absorption and too noisy spectral bands, only 177 spectral bands are used in our experiments. Seven classes of interest are labeled in the image, which are: large buildings, wetlands, residential areas, roads, suburbs, industrial areas and salt soil. A three-band false color image, the training data set and the ground truth are all illustrated in Fig. 4. The number of training and testing samples are shown in Table. III. More information about the data can be found in [13].

[2]Available at http://aviris.jpl.nasa.gov/html/aviris.freedata.html.

B. Parameter Selection

Our model has two hyper-parameters: the regularization parameter λ and the heat kernel parameter σ. The relationship between λ and the overall classification accuracy is described in Fig. 5. From Fig. 5, the performance of SVM/LLPP is stable with respect to λ. It achieves good results when λ varies from 10^{-3} to 10^2. However, it would over smooth the image when λ is too large, $\lambda = 10^3$ for example. Thus, in our experiments, λ is set to be 10 for all the three data sets. The heat kernel parameter σ affects the computation of similarity largely. In our experiments, σ is determined by the average of the variance of each dimension for the locally selected pixel vectors. The computation of σ by local method can improve the similarity distinction among local pixel vectors.

There are also two parameters in SVM, the regularization parameter C and the RBF kernel parameter γ, should be set. These two parameters are chosen by five-fold cross-validation. The value of log form of C is selected from $\{2, 3, ..., 10\}$. The candidate set for log value of γ is $\{-10, -9, ..., -2\}$.

C. Experiments on Center of Pavia Data

Firstly, we test our method on a benchmark data set, the Center of Pavia data. According to III-B, the parameters of SVM for the original pixel-wise classification are set: $C = 2^8, \gamma = 2^{-4}$.

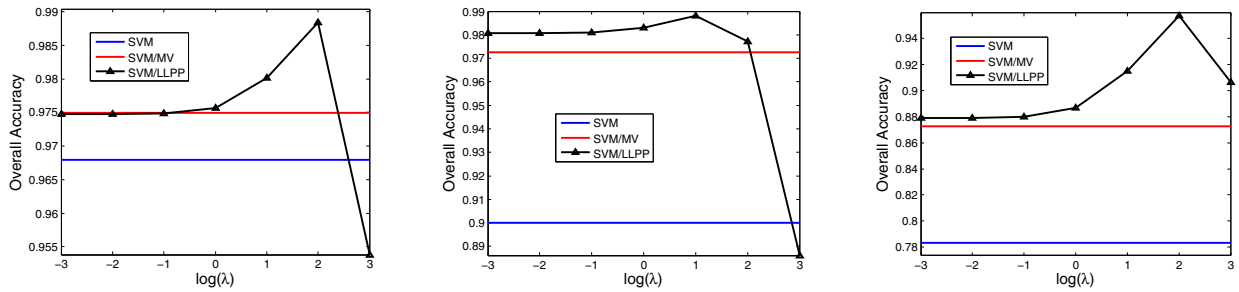
The classification results are listed in Table. I. As can be seen from Table. I, our method performs best on Overall classification Accuracy (OA), Average classification Accuracy (AA) and kappa coefficient (κ) [19] . Compared with SVM/MV, the average classification accuracy of our method is about one percent higher. This indicates that our method can combine spectral information and spatial information effectively.

The classification maps of different methods are shown in Fig. 6. Compared with pixel-wise classification by SVM, the classification map of SVM/MV and SVM/LLPP are more smooth. This benefits from the spatial information. Compared with SVM/MV, some important details, such as thin long roads, are kept well in our method. This is owing to the local label probability propagation in our method, i.e., the label probability propagation just occur among the local spatial neighbors with closer similarity relationship. Some mini land covers those consist of one pixel or only a few pixels are also smoothed away in our method. They are replaced by their relatively similar majority surroundings. The trade-off between the spatial information and spectral information is not an easy job when handling the hyperspectral image classification with complex noise.

D. Experiments on Indian Pines Data

For the Indian pines data, the parameters of SVM for the original pixel-wise classification are set: $C = 2^7, \gamma = 2^{-5}$.

The classification accuracies are shown in Table. II. From Table. II, the global and most of the class-specific classification accuracies of our method are higher than that of the original pixel-wise classification and SVM/MV. This demonstrates the effectiveness of local label probability propagation. The classification maps are depicted in Fig. 7. As can be seen from Fig. 7, the classification results of the SVM/MV and



(a) Center of Pavia image (b) Indian Pines data (c) Moffett Field data

Fig. 5: The performance of SVM/LLPP vs. parameter λ. The performance of SVM/LLPP is stable with respect to λ. It achieves good results when λ varies form 10^{-3} to 10^{2}.

TABLE I: Classification accuracy on Center of Pavia image. Classes of interest, number of training and testing samples, the classification accuracies (%) by different methods are given. In the last three rows, OA is overall accuracy, AA is average accuracy, κ is kappa coefficient [19].

Class	Samples		SVM	SVM/MV	SVM/LLPP
	Train	Test			
Water	745	64533	97.81	97.88	**98.26**
Trees	785	5723	92.89	94.29	**95.35**
Meadows	797	2108	95.11	**96.44**	96.25
Bricks	485	1667	78.22	83.86	**87.46**
Bare Soil	820	5729	94.97	96.82	**97.99**
Asphalt	678	6907	95.70	97.89	**98.67**
Bitumen	808	6479	96.03	98.77	**99.04**
Tiles	223	2899	99.69	**100.00**	**100.00**
Shadows	195	1790	99.75	**99.80**	**99.80**
OA	-	-	96.80	97.50	**98.01**
AA	-	-	94.46	96.19	**96.98**
κ	-	-	94.25	95.50	**96.42**

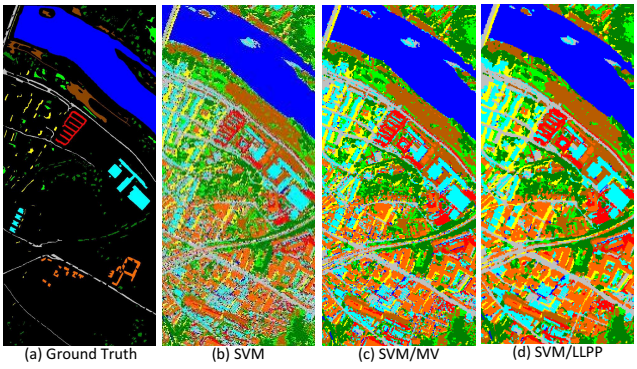


Fig. 6: Classification maps of Center of Pavia image. From left to right: (a) Ground Truth, (b) SVM, (c) SVM/MV, (d) SVM/LLPP.

SVM/LLPP have more continuities for most of the structures. This benefits from the incorporation of the spatial information. In the classification map of our method, most of the small objects with tens of pixels are kept well and the isolated noisy labels are replaced by their similar spatial neighbors at the same time. These illustrate the effectiveness of locally reliable label probability propagation.

TABLE II: Classification accuracy on Indian Pines data. Classes of interest, number of training and testing samples, the classification accuracies (%) by different methods are given. In the last three rows, OA is overall accuracy, AA is average accuracy, κ is kappa coefficient.

Class	Samples		SVM	SVM/MV	SVM/LLPP
	Train	Test			
Corn-no till	357	1071	84.03	95.89	**98.32**
Corn-min till	208	622	85.05	97.43	**98.55**
Grass-pasture	121	362	96.96	**98.34**	96.13
Grass-trees	183	547	97.99	99.45	**100.00**
Hay-windrowed	120	358	99.72	**100.00**	**100.00**
Soybean-no till	243	729	78.88	92.59	**95.47**
Soybean-min till	614	1841	89.08	97.07	**99.78**
Soybean-clean	149	444	89.86	97.75	**99.10**
Woods	317	948	99.47	99.68	**99.89**
OA	-	-	90.00	97.26	**98.80**
AA	-	-	91.23	97.58	**98.58**
κ	-	-	88.24	96.78	**98.59**

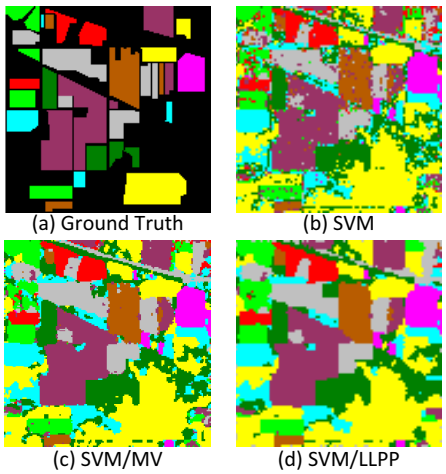


Fig. 7: Classification maps of Indian Pines data. From left to right: (a)Ground Truth, (b)SVM, (c)SVM/MV, (d)SVM/LLPP.

E. Experiments on Moffett Field Data

For the Moffett Field data, parameters of SVM for the original wave-band feature are set: $C = 2^{10}, \gamma = 2^{-3}$.

TABLE III: Classification accuracy on Moffett Field data. Classes of interest, number of training and testing samples, the classification accuracies (%) by different methods are given. In the last three rows, OA is overall accuracy, AA is average accuracy, κ is kappa coefficient.

Class	Samples		SVM	SVM/MV	SVM/LLPP
	Train	Test			
Large buildings	457	838	69.69	79.47	**85.56**
Wetlands	522	567	88.18	91.01	**92.77**
Residential areas	666	8280	63.39	75.16	**82.31**
Roads	646	858	83.57	**89.16**	87.41
Suburbs	911	23877	83.48	91.84	**95.27**
Industrial areas	574	852	88.38	94.84	**97.89**
Salt soil	401	655	62.44	68.40	**73.89**
OA	-	-	78.34	87.27	**91.50**
AA	-	-	77.02	84.27	**87.87**
κ	-	-	60.58	75.54	**83.27**

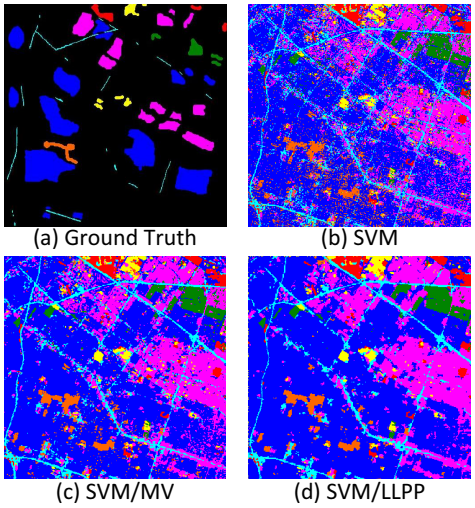


Fig. 8: Classification maps of Moffett Field data. From left to right: (a) Ground Truth, (b) SVM, (c) SVM/MV, (d) SVM/LLPP.

The classification accuracies are shown in Table. III. The SVM/LLPP method performs the best global classification accuracy from Table. III, which is more than 10 percentage points higher than that of the original pixel-wise classification. The classification maps are depicted in Fig. 8, from which the long thin road in the left part of the image is retained well. This illustrates that the SVM/LLPP method can keep long thin objects while smoothing the noisy label image. The reason behind this phenomenon is that the SVM/LLPP method only propagates label to the spatial neighbors that are more similar to the central pixel, but not include the spatial neighbors that have relatively low similarities in the local region.

IV. Conclusion

In this paper, we have constructed a graph Laplacian on the probability estimates by SVM in the hyperspectral image classification. Through seed selection and locally reliable label probability propagation, the spectral information and spatial information are combined together. Moreover, our model can be optimized efficiently by solving a sparse linear equation system. Comparative classification experiments on different hyperspectral images show that our method can achieve accurate classification results. In the future, more sophisticated features, such as Gabor feature, will be explored to improve the performance of the pixel-wise classification in our model, and more reasonable label propagation strategy for hyperspectral image classification will also be explored.

References

[1] X. Jia, "Simplified maximum likelihood classification for hyperspectral data in cluster space," in *IGARSS*, June 2002, pp. 24–28.

[2] J.-M. Yang, P.-T. Yu, and B.-C. Kuo, "A nonparametric feature extraction and its application to nearest neighbor classification for hyperspectral image data," *IEEE Trans. Geoscience and Remote Sensing*, vol. 48, no. 3-1, pp. 1279–1293, 2010.

[3] G. Camps-Valls and L. Bruzzone, "Kernel-based methods for hyperspectral image classification," *IEEE Trans. Geoscience and Remote Sensing*, vol. 43, no. 6, pp. 1351–1362, 2005.

[4] J. A. Benediktsson, J. A. Palmason, and J. R. Sveinsson, "Classification of hyperspectral data from urban areas based on extended morphological profiles," *IEEE Trans. Geoscience and Remote Sensing*, vol. 43, no. 3, pp. 480–491, 2005.

[5] M. Fauvel, J. A. Benediktsson, J. Chanussot, and J. R. Sveinsson, "Spectral and spatial classification of hyperspectral data using svms and morphological profiles," *IEEE Trans. Geoscience and Remote Sensing*, vol. 46, no. 11, pp. 3804–3814, 2008.

[6] Y. Tarabalka, J. A. Benediktsson, and J. Chanussot, "Spectral-spatial classification of hyperspectral image based on partitional clustering techniques," *IEEE Journal of Selected Topics in Signal Processing*, vol. 47, no. 8, pp. 2973–2987, Aug. 2009.

[7] Y. Tarabalka, J. Chanussot, and J. A. Benediktsson, "Segmentation and classification of hyperspectral images using minimum spanning forest grown from automatically selected markers," *IEEE Trans. Systems, Man, and Cybernetics, Part B*, vol. 40, no. 5, pp. 1267–1279, 2010.

[8] Y. Tarabalka, J. A. Benediktsson, J. Chanussot, and J. C. Tilton, "Multiple spectral-spatial classification approach for hyperspectral data," *IEEE Trans. Geoscience and Remote Sensing*, vol. 48, no. 11, pp. 4122–4132, 2010.

[9] K. Bernard, Y. Tarabalka, J. Angulo, J. Chanussot, and J. A. Benediktsson, "Spectral-spatial classification of hyperspectral data based on a stochastic minimum spanning forest approach," *IEEE Trans. Image Processing*, vol. 21, no. 4, pp. 2008 – 2021, 2012.

[10] B. Zhang, S. Li, X. Jia, L. Gao, and M. Peng, "Adaptive markov random field approach for classification of hyperspectral imagery," *IEEE Geosci. Remote Sensing Lett.*, vol. 8, no. 5, pp. 973–977, 2011.

[11] X.Jia and J.A.Richards, "Managing the spectral-spatial mix in context classification using markov random fields," *IEEE Geosci. Remote Sensing Lett.*, vol. 5, no. 2, pp. 311–314, 2008.

[12] G. Camps-Valls, T. V. B. Marsheva, and D. Zhou, "Semi-supervised graph-based hyperspectral image classification," *IEEE Trans. Geoscience and Remote Sensing*, vol. 45, no. 10, pp. 3044–3054, 2007.

[13] J. Bai, S. Xiang, and C. Pan, "A graph-based classification method for hyperspectral images," *IEEE Trans. Geoscience and Remote Sensing*, vol. 51, no. 2, pp. 803–817, 2013.

[14] T.-F. Wu, C.-J. Lin, and R. C. Weng, "Probability estimates for multi-class classification by pairwise coupling," *Journal of Machine Learning Research*, vol. 5, pp. 975–1005, 2004.

[15] C. Chih-Chung and L. Chih-Jen, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 1–27, 2011.

[16] D. Cai, X. He, J. Han, and T. S. Huang, "Graph regularized nonnegative matrix factorization for data representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 8, pp. 1548–1560, 2011.

[17] F. Chung, *Spectral Graph Theory*. AMS, 1997.

[18] U. von Luxburg, "A tutorial on spectral clustering," *Statistics and Computing*, vol. 17, no. 4, pp. 395–416, 2007.

[19] J. A.Richards and X. Jia, *Remote Sensing Digital Image Analysis*. Springer, 2006.