Australian Journal of Basic and Applied Sciences, 5(8): 374-382, 2011
ISSN 1991-8178

Contextual Classification of Hyperspectral Remote Sensing Images Using SVM-PLR

[1]Mahdi Khodadadzadeh, [2]Hassan Ghassemian

[1,2]Faculty of Electrical and Computer Engineering, Tarbiat Modares University, Tehran, Iran.

Abstract: In this paper, we propose a novel contextual classification of hyperspectral data. We use probabilistic label relaxation (PLR) process to incorporate context information into the spectral pixelwise classification procedure. In conventional PLR procedure, first a maximum likelihood classification is performed and class probabilities are computed by using multivariate normal models. However this method is not efficient for hyperspectral data with limited training samples. In this paper we suggest to use support vector machine (SVM) in order to initial classification and also use class probability estimates which are obtained from SVM classification for PLR postprocess. We call this proposed method as SVM-PLR. Experimental results are presented for two types of hyperspectral images, agricultural and urban data. The proposed method improves dramatically classification accuracies, when compare to spectral pixelwise classification. Moreover our proposed method can improve performance of conventional PLR postprocess for hyperspectral data.

Key words: Hyperspectral images, contextual classification, probabilistic label relaxation (PLR), support vector machine (SVM).

INTRODUCTION

Hyperspectral imagery records a detailed spectrum of light arriving at each pixel (C.-I. Chang, 2007). The rich spectral information of hyperspectral data has made it possible to distinguish materials and land cover classes present on the Earth's surfaces. One of the most important tasks for hyperspectral images is classification which is a method by which labels may be attached to pixels in view of their spectral character (J. A. Richards and X. Jia, 2006). More than 100 spectral channels in hyperspectral images improve overall image classification accuracy; however this number of spectral channels presents some constraints. One of the most important constraints is limited training samples of hyperspectral images. While it is often stated that the number of training samples for each class should comprise at least 10-30 times the number of wavebands, this quantity is precious and unavailable for hyperspectral data (Foody and Mathur, 2006).

Various classification technique have been proposed that label a pixel on the basis of its spectral properties alone, with no attention to labels which are assigned to neighboring pixels. However, in remote sensing images the pixels in the neighborhood are likely to assign the same labels, because imaging sensors acquire significant portions of energy from adjacent pixels and because ground cover types generally occur over a region that is large compared with the size of a pixel (J. A. Richards and X. Jia, 2006). Therefore integration of contextual and spectral information can mitigate the problem of insufficient number of training samples. Landgrebe and his research group were pioneers in using spatial information for improving remote sensing image classification (R. Kettig and D. Landgrebe, 1976). During recent years, various approaches have been discussed to incorporate contextual information (Q. Jackson and D. Landgrebe, 2002; P. Zhong and R. Wang, 2007; A. Plaza, J. A. Benediktsson, 2009; Y. Tarabalka and J. A. Benediktsson, 2009; M. Khodadadzadeh and H. Ghassemian, 2010).

One category of these approaches is to perform postprocess after the image has been classified by a pixel-wise classifier. The postprocessing filters and post-classification approaches like Probabilistic Label Relaxation (PLR) which use fixed-window for incorporating contextual information are in this category.

In conventional PLR procedure, first a maximum likelihood classification is performed and class probabilities are computed by using multivariate normal models. However, the effectiveness of maximum likelihood classification depends upon reasonably accurate estimation of the mean vector and the covariance matrix for each spectral class. Consistent estimation is dependent upon having a sufficient number of training samples that is not available for hyperspectral images. It is also significant to note that classification time, increases quadratically with number of spectral components for the maximum likelihood classifier (J. A.

Corresponding Author: Hassan Ghassemian, Faculty of Electrical and Computer Engineering, Tarbiat Modares University, Tehran, Iran
E-mail: ghassemi@modares.ac.ir.

Richards and X. Jia, 2006).

SVMs have shown good performance in hyperspectral data classification particularly in small training samples size. SVM is originally a binary classifier while in remote sensing application we face to several classes of interest. Moreover SVMs produce a value that is not a probability. Various approaches have been proposed to alleviate these problems (T.-F. Wu, C.-J. Lin and R. C. Weng, 2004; M. Gonen and A. G. Tanugur, 2008; A. Mathur and G. M. Foody, 2008).

In this paper we propose to use a precious technique that provide multi-class probability estimates for SVM classifier and then by using these class probability estimates, perform PLR after SVM classification to incorporate context. Using PLR postprocessing method based on SVM classifier (SVM-PLR), has shown better performance in comparison with traditional ML-PLR classification scheme. The block diagram of the proposed method is shown in Fig. 1.

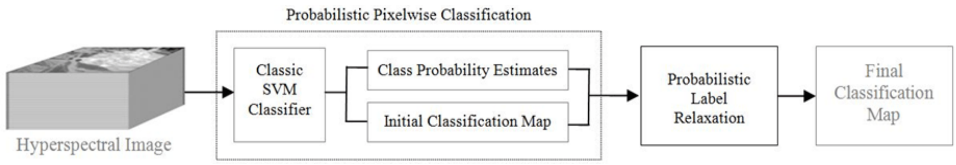

Fig. 1: Block diagram of the proposed approach

Two hyperspectral airborne images were used to demonstrate experimental results: a 103-band Reflective Optics System Imaging Spectrometer (ROSIS) image of the University of Pavia, Italy, and a 220-band AVIRIS image taken over the Northwestern Indiana's Indian Pine site. Although the proposed method has been tested on these hyperspectral images, the proposed method is general and can be applied for other types of data as well.

Probabilistic Pixelwise Classification:

The first step is to perform a probabilistic pixelwise classification of the hyperspectral image. We propose to use a SVM classifier and then compute probability estimates for multiclass classification. SVMs search for optimal separating hyperplane with maximal distance between the hyperplane and the nearest samples (support vectors) from each of the two classes. We briefly describe SVM algorithm for a supervised binary classification problem as following.

Let us assume that we have *N* training vectors from *d* dimensional feature space $\mathbf{x}_i \in \Re^d \; (i = 1, 2, \ldots, N)$ which labeled with $y_i \in \{-1, +1\}$ Function $\hat{y} = \operatorname{sgn}[f(\mathbf{x})]$ is used for making decision on input feature vector x, where *f*(x) is discriminant function of SVM which is given as

$$f(\mathbf{x}) = \mathbf{w}.\mathbf{x} + b, \tag{1}$$

where w is weight vector perpendicular to separating hyperplane and $b \in \Re$ is a bias. These values can specify a hyperplane and can be calculated by following optimization problem

$$\begin{cases} \text{minimize:} \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{N} \xi_i \\ \text{subjected to:} \; y_i(\mathbf{w}.\mathbf{x} + b) \geq 1 - \xi_i \\ \xi_i \geq 0, \; \forall i = 1, 2, \ldots, N, \end{cases} \tag{2}$$

hear, *C* is the penalty parameter which control the trade-off between training accuracy and generalization and ξ_i 's are *slack variables* that introduced in order to deal with misclassified vectors. The optimization problem (2) can be solved using Lagrange multipliers and then becomes

$$\begin{cases} \text{minimize: } \sum_{i=1}^{N} \alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j (\mathbf{x}_i . \mathbf{x}_j) \\ \text{subjected to: } \sum_{j=1}^{N} \alpha_i y_i = 0 \text{ and } 0 \le \alpha_i \le C, \\ \forall i = 1, 2, \ldots, N. \end{cases} \quad (3)$$

By replacing inner products $(\mathbf{x}_i . \mathbf{x}_j)$ with a kernel function that fulfill Mercer's condition like Gaussian radial basis function (4), the final discriminant function $f(\mathbf{x})$ is defined as (5), which is belonging to the category of nonlinear discriminant function.

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp(\gamma \left\| \mathbf{x}_i - \mathbf{x}_j \right\|^2) \quad (4)$$

$$f(\mathbf{x}) = \sum_i \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) + b. \quad (5)$$

In (J. Platt, 1999) two precious techniques are proposed for computing multi-class probability estimates based on combining all pairwise comparisons. We use one which is described as following.

In this technique we want to estimate, for each pixel $\mathbf{x}$, the probabilities of belonging to each K class as

$$\mathbf{p} = \left\{ p_i = p(\omega_i \mid \mathbf{x}), i = 1, \ldots, K \right\}. \quad (6)$$

The probabilities are computed by solving the following optimization problem that has a unique solution and can be solved by a linear system

$$\begin{aligned} &\mathbf{p}^* = \min_{\mathbf{p}} \sum_{i=1}^{K} \sum_{j:j \ne i} \left(r_{ji} p_i - r_{ij} p_j \right)^2 \\ &\text{subjected to } \sum_{i=1}^{K} p_i = 1, \; p_i \ge 0, \forall i, \end{aligned} \quad (7)$$

where $\mathbf{p}^*$ is the set of probability estimates and r_{ij} are pairwise class probabilities that $r_{ij} + r_{ji} = 1, \forall i \ne j.$

It is clear that for approximating pairwise class probabilities, subset of training samples of total that belongs to i and j classes, $\mathbf{x}_k \in \Re^d \; (i = 1, 2, \ldots, L)$ is considered, which labeled by $y_i \in \left\{-1, +1\right\}$. Platt's algorithm is used to solve this two class problem which is proposed to approximate a posterior class probability by a sigmoid function

$$r_{ij} \equiv \Pr(y = 1 \mid \mathbf{x}) \approx \frac{1}{1 + e^{Af + B}}, \quad (8)$$

where $f = f(\mathrm{x})$ is a decision value that produced by SVM classifier (5). Let for each training sample $p_k = \Pr(y = 1 \mid \mathbf{x}_k)$ then, by minimizing the following negative log-likelihood function the best parameters (A^*, B^*) can be determined as

$$(A^*, B^*) = \min_{(A,B)} \left[-\sum_{k=1}^{L} \left(t_k \log(p_k) + (1 - t_k) \log(1 - p_k) \right) \right], \quad (9)$$

where t_k is the MAP estimate for target probability that can computed as

$$t_k = \begin{cases} \dfrac{N_+ + 1}{N_+ + 2} & \text{if } y_k = 1 \\ \dfrac{1}{N_- + 2} & \text{if } y_k = -1 \end{cases}, \quad \text{for } k = 1, 2, \ldots, L, \quad (10)$$

where N_+ and N_- are number of y_k's positive and negative respectively. An improved implementation of Platt's algorithm is presented in (H. Lin and C. Lin, 2003) which is theoretically converges and avoids numerical difficulties.

Probabilistic Label Relaxation:

Probabilistic Label Relaxation (PLR) is a postprocess procedure which is used fixed-window to incorporate contextual information and assign the most appropriate class for each pixel. We describe PLR algorithm as following.

Let the set of probabilities for a pixel (m) represented by

$$p_m(\omega_i) \quad i = 1, 2, \ldots, K, \quad \text{where} \sum_{i=1}^{K} p_m(\omega_i) = 1.$$

Suppose that a neighborhood Nm is defined surrounding pixel m as shown in Fig. 2. At kth iteration, the correct set of label probabilities for the pixel m that having taken account both of spectral and contextual information is defined as

$$p_m^{k+1}(\omega_i) = \frac{p_m^k(\omega_i) Q_m^k(\omega_i)}{\sum_{i=1}^{K} p_m^k(\omega_i) Q_m^k(\omega_i)},$$

here $Q_m^k(\omega_i)$ is neighborhood function at kth iteration that can be defined as

$$Q_m^k(\omega_i) = \sum_{n \in N_m} d_n \sum_{j=1}^{K} p_{mn}(\omega_i \mid \omega_j) p_n^k(\omega_j),$$

where d_n s are neighbor weights that recognize that some neighbors may be more influential than others and $p_{mn}(\omega_i \mid \omega_j)$ is the capability measure that can be estimated directly from initially classification map.

Unless there is good reason to do otherwise the neighbor weights are generally chosen all to be the same (J. A. Richards and X. Jia, 2006). Equation (12) is applied as many times as necessary to ensure that $p_m(\omega_i)$ have stabilized.

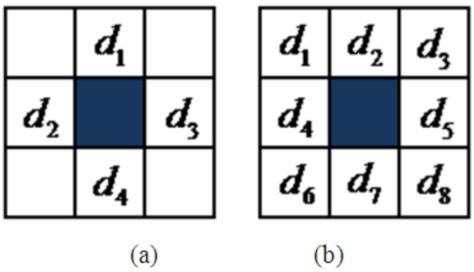

Fig. 2: Definition of neighborhood with assigned neighbor weights. (a) four neighborhood system, (b) eight neighborhood system

Experimental Results:

We use two types of hyperspectral airborne images to demonstrate experimental results, agricultural and urban, which are Indiana image and University of Pavia image respectively.

Indiana Image:

The hyperspectral data used in the first experiment was recorded by AVIRIS sensor over the Indiana Pines test site in Northwestern Indina. The image has spatial dimension of 145 145 pixels, and the spatial resolution is 20 m per pixel. This data has originally 220 spectral bands, however due to the effect of atmospheric phenomena, 20 spectral bands (104-108, 150-163, and 220) were discarded. A three-band false color image and the available groundtruth map data are shown in Fig. 3(a) and Fig. 3(b) respectively.

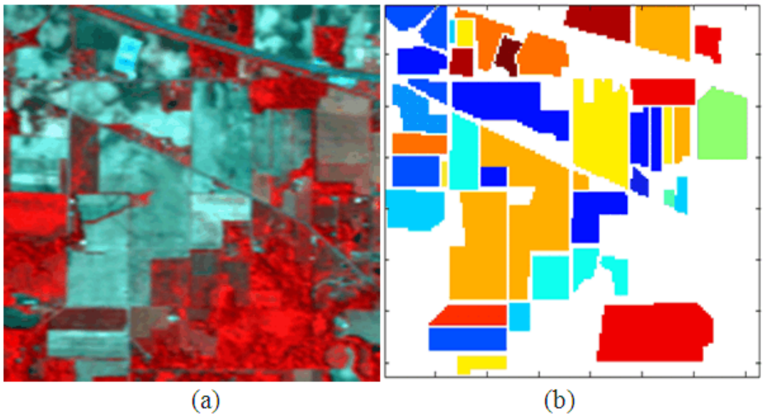

Fig. 3: Indiana image. (a) Three-band color composite (b) Reference Map

The available original groundtruth map data contain sixteen classes of different types of crops, however six classes (alfalfa, corn, grass/pasture-mowed, oats, bldg-grass-tree-drives, stone-steel towers) that contain a small number of samples (less than 380 samples) in the groundtruth map are discarded. The remaining classes with the numbers of test and training samples for each class are detailed in Table. I. The training set is chosen randomly from groundtruth image and the remaining samples are considered as the test set. First as shown in Fig. 4 initial SVM classification was performed using the multiclass pairwise (one versus one) SVM classifier, with the Gaussian RBF kernel. The parameters C and γ were determined by fivefold cross validation, which gave C = 1024 and $\gamma = 2^{-7}$. Then the probabilities estimates were obtained using equation (7) and based on them, the final classification map (see Fig. 4(d)) was obtained after postprocessing using proposed PLR method. Table. I show the class-specific accuracies (that is the percentage of correctly classified pixels for each class) for the ten different land-cover classes as well as Overall accuracy (OA), average accuracy (AA) and kappa coefficient (K). We achieve 89.34% OA, 91.32% AA, and 87.69% K value in this data; where these values are 77.32%, 82.07%, and 74.09% for SVM classification, respectively. The classical PLR algorithm using the multivariate normal models to compute class probabilities, in order to incorporate spatial context into a preexisting ML classification. Therefore a feature reduction algorithm is applied to reduce the spectral dimension of pixel vectors in this data. Because large dimensionality of feature vectors in hyperspectral data may cause the problem of the covariance matrix singularity or inaccurate parameter estimation results in ML classification. We proposed to use PCFA (A. Jensen and A. Solberg, 2007) method to reduce the 200 spectral bands of this data to ten. In experiments this method showed better performance than the other common methods like PCA and ICA for feature reduction in this data. As detailed in Table. I, ML classification shows low classification accuracies, for example *Soybeans-min till* class which describes mostly large regions in the image, shows classification accuracy 42.87 which is not acceptable. After PLR process on this ML classification map, the classification accuracies are improved in a range of 0.2%–25.57%. Although integration of spatial and spectral information increases classification accuracies in the most of the classes, these values are still not proper for classes like *Corn-min till*, *Soybeans-min till* and *Woods*. By comparing classification accuracies resulted from our proposed PLR method to traditional PLR method which is based on ML classification, we considered that for those mentioned classes with low classification accuracies, these values were increased dramatically when the class probabilities obtained from SVM classification are used in the PLR procedure. In particular, the class *Soybeans-min till* is much more accurately classified when the proposed method is used (improvement of the classification accuracy by 27.96%). The final classification maps are shown in Fig. 5. The use of contextual information by performing PLR process significantly reduces noise in the classification map which leads to improved classification accuracies. Good classification results in this agricultural hyperspectral data shows that, the proposed approach is particularly suitable for classification of images containing large spatial structures.

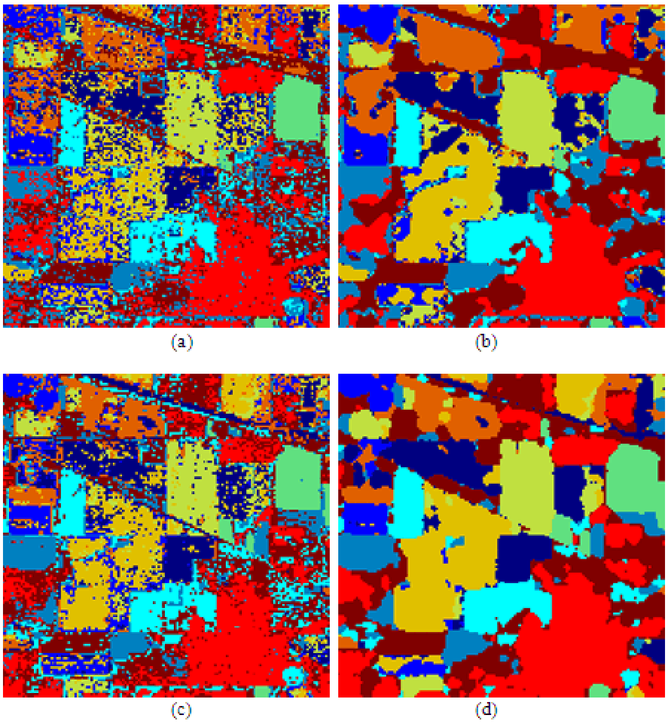

Fig. 4: Indiana image, the result of (a) ML, (b) ML-PLR, (c) SVM and (d) SVM-PLR classification

Table 1: Classes, number of samples, and class specific accuracies in percentage for the Indiana image

Class		No. of Samples		10 spectral bands		200 spectral bands	
No	Name	Test	Train	ML	ML+PLR	SVM	SVM+PLR
1	Corn-no till	1384	50	58.51	72.38	68.20	79.64
2	Corn-min till	784	50	60.67	69.18	70.50	78.06
3	Grass/pasture	447	50	74.04	86.92	96.78	99.20
4	Corn	697	50	85.68	97.46	91.16	98.80
5	Hay-windrowed	439	50	99.39	99.59	99.18	100.00
6	Soybeans-no till	918	50	73.24	82.75	82.02	95.35
7	Soybeans-min till	2418	50	42.87	55.63	63.53	83.59
8	Soybeans-clean till	564	50	67.59	93.16	86.16	96.42
9	Wheat	162	50	91.19	98.15	87.33	99.00
10	Woods	1244	50	52.37	62.37	75.79	83.16
Overall Accuracy (OA)		-	-	65.81	78.05	77.35	89.34
Average Accuracy (AA)				70.55	82.65	82.07	91.32
Kappa (K)				61.09	74.94	74.09	87.69

University of Pavia Image:

The *University of Pavia* image was recorded by the ROSIS optical sensor over the urban area of the University of Pavia. The original spatial size of image is 610 × 340 pixels with a spatial resolution of 1.3 m/pixel Here we are interested only in a part of this image (340 × 340) which contains all nine classes of interest (*Asphalt*, *Meadows*, *Gravel*, *Trees*, *Meta sheets*, *Bare soil*, *Bitumen*, *Bricks*, *Shadows*). A three-band false color image and the groundtruth map data are shown in Fig. 5(a) and Fig. 5(b) respectively.

The training set is chosen randomly from groundtruth image and the remaining samples are considered as the test set. The original image has 115 spectral bands however 12 most noisy bands were discarded and as Indiana image, for comparing our proposed PLR method to traditional PLR, we reduce the remaining spectral bands to ten by using the method of PCFA (A. Jensen and A. Solberg, 2007).

Multiclass, one versus one, SVM classification was performed on the original image using the Gaussian RBF kernel. The parameters C and γ were determined by fivefold cross validation, which gave $C = 128$ and $\gamma = 0.125$ for this data. After the conventional SVM pixelwise classification, the probabilities estimates were obtained using equation (7). Then the final classification map was obtained after postprocessing using proposed PLR method. Table. II give the class-specific and the global classification accuracies, respectively, for the ML and SVM classification, without and with the PLR step. The classification maps for the pixel wise and the contextual classification after the PLR are shown in Fig. 6.

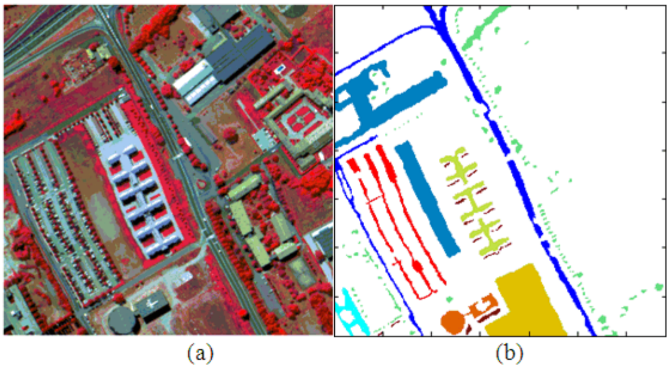

Fig. 5: Part of the university of pavia image. (a) Three-band color composite (b) Reference Map

The SVM classifier correctly classified 93.91% of pixels from the test set while this value is 82.28% for ML classification. The best classification accuracies are achieved for the classes *Trees*, *Meta sheets* and *Shadows*. About classes *Meadows* and *Bare soil*, mostly large regions in the image, the accuracies are improved by 13.32% and 25.99% respectively, compared to the pixelwise ML classification. The accuracies were considerably improved after PLR postprocess, with the improvement being more significant when this step is performed after a pixel wise SVM classification.

As detailed in Table. II the proposed contextual classification improves the classification accuracies for almost all the classes, except for the class *Trees*. For this class, the difference is not significant and the accuracy is high enough. For the other classes, classification accuracies are improved in a range of 0.96%–21.68%. The best global accuracies are obtained when performing our proposed PLR procedure. We achieve 98.58% OA, 98.14% AA, and 98.28% K value in this data; Where these values are 93.91%, 94.06%, and 92.65% for SVM classification, respectively.

As shown in Fig. 6, the use of contextual information by performing PLR process significantly reduces noise in the classification map. Specially in the case of this data (that is a urban hyperspectral image comprised of complex boundaries), these significant classification results, show that, the proposed approach is also suitable for classification of images containing small spatial structures.

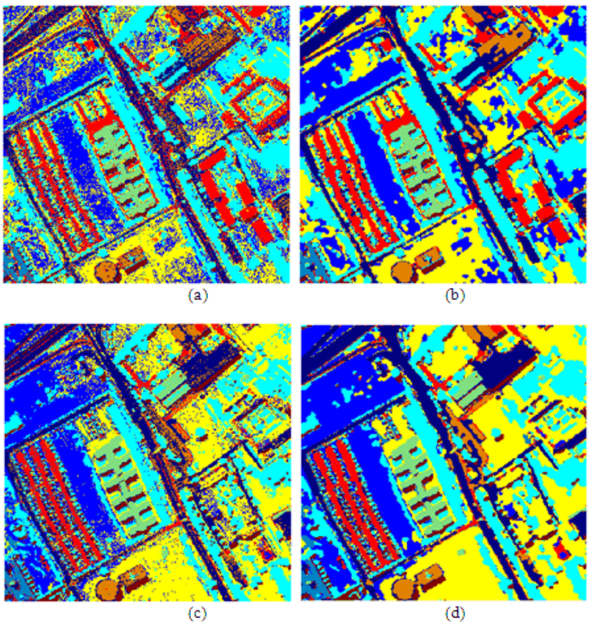

Fig. 6: University of Pavia image, the result of (a) ML, (b) ML-PLR, (c) SVM and (d) SVM-PLR classification

Table 2: Classes, number of samples, and class specific accuracies in percentage for the university of pavia image

Class		No. of Samples		10 spectral bands		200 spectral bands	
No	Name	Test	Train	ML	ML+PLR	SVM	SVM+PLR
1	Asphalt	3380	50	71.66	92.16	87.99	97.81
2	Meadows	5166	50	83.57	98.27	96.89	99.69
3	Gravel	520	50	83.68	93.51	89.65	95.44
4	Trees	1626	50	97.85	99.58	98.75	98.99
5	Meta sheets	1259	50	95.69	98.74	99.33	99.70
6	Bare soil	3668	50	67.37	76.52	93.36	98.20
7	Bitumen	708	50	84.04	92.22	90.24	95.38
8	Bricks	1659	50	79.81	93.97	90.64	98.36
9	Shadows	320	50	97.30	98.11	99.73	99.73
Overall Accuracy (OA)		-	-	82.28	92.31	93.91	98.58
Average Accuracy (AA)				84.55	93.67	94.06	98.14
Kappa (K)				76.27	90.66	92.65	98.27

Conclusion:

In this paper, we propose a novel contextual classification of hyperspectral data. We use probabilistic label relaxation (PLR) process to incorporate context information into the spectral pixelwise classification procedure. In conventional PLR procedure, first a maximum likelihood classification is performed and class probabilities are computed by using multivariate normal models. However this method is not efficient for hyperspectral data with limited training samples. In this paper we suggest to use support vector machine (SVM) in order to initial classification and also use class probability estimates which are obtained from SVM classification for PLR postprocess.

Experimental results are presented for two types of hyperspectral images, agricultural image (that is an image including large spatial structures comprised of flat areas) and urban data (that is a image containing small spatial structures comprised of complex boundaries). The proposed method improves dramatically classification accuracies in both images and provides classification maps with more homogeneous regions, when compare to spectral pixelwise classification. Moreover our proposed method can improve performance of conventional PLR postprocess for hyperspectral data.

ACKNOWLEDGEMENTS

The authors would like to thank Prof. Benediktsson of the University of Iceland, Iceland and Prof. Gamba of the Univeristy of Pavia, Italy, for providing the Pavia data set. This work was supported in part by the Iran Research Institute for ICT-ITRC under contract No. 500/17581.

REFERENCES

Chang, C.-I., 2007. Hyperspectral Data Exploitation: Theory and Applications, Wiley-Interscience.

Foody, G.M. and A. Mathur, 2006. The use of small training sets containing mixed pixels for accurate hard image classification: training on mixed spectral responses for classification by a SVM, Remote Sens. Environ., 103(7): 179-189.

Ghassemian, H. and D. Landgrebe, 1988. Object-oriented feature extraction method for image data compaction, IEEE Control System Magazine, 8(3): 42-46.

Gonen, M., A.G. Tanugur and E. Alpaydm, 2008. Multiclass posterior probability support vector machines, IEEE Trans. Neural Networks, 19(1): 130-139.

Jackson, Q. and D. Landgrebe, 2002. Adaptive bayesian contextual classification based on markov random fields, IEEE Trans. Geosci. Remote Sens., 40(11): 2454-2463.

Jensen, A. and A. Solberg, 2007. Fast hyperspectral feature reduction using piecewise constant function approximations, IEEE Geoscience and Remote Sensing Letters, 4(4): 547-551.

Kettig, R. and D. Landgrebe, 1976. Classification of multispectral image data by extraction and classification of homogenous objects, IEEE Trans. Geosci. Remote Sens., 14: 19-26.

Khodadadzadeh, M., R. Rajabi and H. Ghassemian, 2010. Combination of region-based and pixel-based hyperspectral image classification using erosion technique and MRF model, In Proc. 18th Iranian Conf. on Electrical Engineering , pp. 294-299, Isfahan, Iran.

Lin, H., C. Lin and R.C. Weng, 2003. A note on plats probabilistic outputs for support vector machines, Dept. Comput. Sci., Nat. Taiwan Univ., Taipei, Taiwan.

Mathur, A. and G.M. Foody, 2008. Multiclass and binary SVM classification: Implications for training and classification users, IEEE Geoscience and Remote Sensing Letters, 5(5): 241-245.

Platt, J., 1999. Probabilistic outputs for support vector machines and comparison to regularized likelihood methods, in Advances in Large Margin Classifiers, MA: MIT Press, 61-74.

Plaza, A., J.A. Benediktsson, J. Boardman, J. Brazile, L. Bruzzone, G. Camps-Valls, J. Chanussot, M. Fauvel, P. Gamba, J. A. Gualtieri, M. Marconcini, J.C. Tilton and G. Trianni, 2009, Recent advances in techniques for hyperspectral image processing, Remote Sens. Environ., 113: 110-122.

Richards, J.A. and X. Jia, 2006. Remote Sensing Digital Image Analysis, Springer-Verlag Berlin Heidelberg.

Tarabalka, Y. and J. A. Benediktsson, 2009. Spectral-spatial classification of hyperspectral imagery based on partitional clustering techniques, IEEE Trans. Geosci. Remote Sens., 47(8): 761-776.

Wu, T.-F., C.-J. Lin and R.C. Weng, 2004. Probability estimates for multi-class classification by pairwise coupling, J. Mach. Learn. Res., 5: 975-1005.

Zhong P. and R. Wang, 2007. A multiple conditional random fields ensemble model for urban area detection in remote sensing optical images, IEEE Trans. Geosci. Remote Sens., 45(12): 3978-3988.