Supervised Composite Kernel Locality Preserving Projection Feature Extraction for Hyperspectral Image Classification

J.H. Wen
School of Science, Northwestern Polytechnical University
Xi'an, China
wjhgdsx@nwpu.edu.cn

W.D. Yan
School of Science, Northwestern Polytechnical University
Xi'an, China
yanweidong@nwpu.edu.cn

Abstract—**Locally preserving projection (LPP) does not take advantage of the spatial correlation of pixels in the image, and the pixels are considered as independent pieces of information. In this paper, a kernel based manifold learning feature extraction method which considers spatial relationship of neighboring pixels, called supervised composite kernel locality preserving projection (SCKLPP), is proposed for hyperspectral image feature extraction. The spatial information and spectral information from original hyperspectral image are combined using composite kernels weight matrix. The nearest neighbor graph is created with the prior class-label information of samples. Experimental results on AVIRIS data set show that the SCKLPP can not only efficiently reduce the dimensionality but also achieve higher accuracies. In addition, the proposed method opens a new field for future developments in which spatial information can be easily integrated into the feature extraction stage.**

Keywords-feature extraction; dimensionality reduction; locally preserving projection; composite kernel; hyperspectral image classification

I. INTRODUCTION

The high number of spectral channels and the relatively small number of labelled training samples of hyperspectral image present a challenge to traditional data processing techniques. Jimenez [1] pointed out that the hyperspectral data is centralized in low-dimensional space because of the high-dimensional space of hyperspectral image is relatively empty. Therefore, reducing the dimensionality of hyperspectral data without loosing important information about objects of interest is a very important issue for the remote sensing community. A number of manifold learning methods have been developed to mitigate the effects of dimensionality on information extraction from hyperspectral data, such as locally linear embedding (LLE) [2-4] and its regularized version of LLE [5-7]. Because it remains a difficult issue to map a new test sample to the low dimensional space, the above LLE and its regularized version of LLE algorithms cannot be easily extended for classification problems. To address this problem, locally preserving projection (LPP) [8] is proposed to be directly applicable for dealing with new test data. LPP efficiently preserve the local structure of data. However, a commom inherent limitation is still exsisted: the original LPP algorithm does not take advantage of the spatial correlation of pixels in the image, and the pixels are considered as a point in the underlying high-dimensional manifold embedded in R^D independently. Hyperspectral data are images; hence, the pixel vectors are also spatially related, and this should be considered [5]. Furthermore, LPP often fails to deliver good performance when data sets are subject to complex nonlinear variations, for it is a linear algorithm in nature. In hyperspectral remote sensing, nonlinear properties are intrinsic, which originate from the multi-scattering between photons and ground targets, within-pixel spectral mixing, and scene heterogeneity [2]. In order to preserve discriminant and nonlinear information in subspace, Cheng redefined the weight matrix and proposed supervised kernel locality preserving projections algorithm (SKLPP) [9]. But they do not take into account the spatial correlation of pixels in the image.

Recent studies show that the exploitation of spatial information is necessary for subspace extraction and classification of hyperspectral imagery [5, 10-11]. Integration of the spatial information for hyperspectral image classification has caused extensive concern [1, 10-13]. In [12], spatial information is extracted using local windows and incorporated in a composite kernel. Velasco-Forevo [13] proposed a strategy for combining information from original and spatially smoothed hyperspectral images using composite kernels for graph-based semisupervised classification. However, as far as we know, very few people have integrated the spatial information into feature extraction for hyperspectral images [5, 15].

Based on [9] and [12], we develop a kernel based manifold learning feature extraction method which consider spatial relationship of neighboring pixels, named supervised composite kernel locality preserving projection (SCKLPP), for hyperspectral image feature extraction in this paper. In the proposed SCKLPP method, the spatial information and spectral information from original hyperspectral image are combined using composite kernels weight matrix. The local geometric relations of the within-class samples in nonlinear kernel feature space are preserved.

The rest of the paper is organized as follows. The locally preserving projection algorithm and spatial-spectral composite kernels are briefly described in Section 2 and 3, respectively. The SCKLPP algorithm is described in Section 4. Section 5 tests the SCKLPP algorithm on hyperspectral data sets. Section 6 ends the paper by presenting some conclusions.

II. LOCALITY PRESERVING PROJECTION

LPP is a linear approximation of Laplacian Eigenmaps [14]. Given a set of *m* training samples $X=[x_1,x_2,\cdots,x_m]$ in R^n. The linear transformation *P* can be obtained by minimizing the following objective function [8]:

$$\min_P \sum_{ij} \left\| y_i - y_j \right\|^2 W_{ij} \tag{1}$$

where $y_i = P^T x_i$. The weight matrix *W* is constructed through the nearest-neighbor graph,

$$W_{ij} = \exp(-\left\| x_i - x_j \right\|^2 / t) \tag{2}$$

where parameter *t* is a suitable constant. Otherwise, W_{ij}=0. For more details of LPP and weight matrix, please refer to [8]. This minimization problem can be converted to solving a generalized eigenvalue problem as follows:

$$XLX^T P = \lambda XDX^T P \tag{3}$$

where $D_{ii} = \sum_j W_{ij}$ is a diagonal matrix, and *L=D-W* is the Laplacian matrix of graph.

III. SPATIAL-SPECTRAL COMPOSITE KERNELS

These spatial-spectral composite kernels were proposed by [12]. It defined the spatial information that a pixel entity x_i is redefined simultaneously both in the spectral domain using its spectral content $x_i^w \in R^{N_w}$ and in the spatial domain by applying some feature extraction to its surrounding area $x_i^s \in R^{N_s}$ which yields N_s spatial (contextual) features, e.g., the mean per spectral band. It constructed a composite kernel with both the spatial (x_i^s) and spectral (x_i^w) information by

$$K(x_i,x_j) = \mu K_s(x_i^s,x_j^s) + (1-\mu) K_w(x_i^w,x_j^w) \tag{4}$$

where $\mu (0<\mu<1)$ is a free parameter that is tuned during the training process, which present a tradeoff between spectral and spatial kernels.

IV. SUPERVISED COMPOSITE KERNEL LOCALITY PRESERVING PROJECTION (SCKLPP)

LPP is a linear method in nature, and it is inadequate to represent the nonlinear hyperspectral data. In hyperspectral remote sensing, nonlinear properties are intrinsic, which originate from the multi-scattering between photons and ground targets, within-pixel spectral mixing, and scene heterogeneity [2]. Moreover, LPP does not take advantage of the spatial correlation of pixels in the image, and the pixels are considered as independent pieces of information. Hyperspectral data are images, hence, the pixel vectors are also spatially related, and this should be considered [5]. By integrating the spatial information into hyperspectral subspace extraction process, higher classification accuracy can be expected.

In SKLPP [9] algorithm, only the spectral information is emphasized. In this paper, we propose a nonlinear manifold learning, named supervised composite kernel locality preserving projection (SCKLPP), which exploits both spectral and spatial information in feature extraction of hyperspectral image. The spatial information and spectral information from original hyperspectral image are combined using composite kernels weight.

Similar to SKLPP algorithm, the objective function of our method is defined as follows:

$$\min_{P^\phi} \sum_{ij} \left\| y_i^\phi - y_j^\phi \right\|^2 W_{ij}^\phi \tag{5}$$

where $y_i^\phi = P^{\phi T}\phi(x_i)$ is the projection of $\phi(x_i)$ onto P^ϕ, $\phi(x_i)$ is the nonlinear mapping of sample x_i, P^ϕ is the transformation matrix.

The objective function (5) can be simplified as

$$\begin{aligned} \min_{P^\phi} \sum_{ij} \left\| y_i^\phi - y_j^\phi \right\|^2 W_{ij}^\phi &= \min_A \sum_{ij} \left\| P^{\phi T}\phi(x_i) - P^{\phi T}\phi(x_j) \right\|^2 W_{ij}^\phi \\ &= 2P^{\phi T} X^\phi (D^\phi - W^\phi) X^{\phi T} P^\phi \\ &= 2P^{\phi T} X^\phi L^\phi X^{\phi T} P^\phi \end{aligned} \tag{6}$$

where $D_{ii}^\phi = \sum_j W_{ij}^\phi$ is a diagonal matrix, $L^\phi = D^\phi - W^\phi$, W_{ij}^ϕ is the weight matrix between any two data point in feature space $F: X^\phi = [\phi(x_1),\phi(x_2),\cdots,\phi(x_m)]$, and it is defined with class information as follows:

$$W_{ij}^\phi = \begin{cases} \exp(-\left\|\phi(x_i)-\phi(x_j)\right\|^2 / t) & \text{if } x_i \text{ and } x_j \text{ belong to the same class} \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

Because $\phi(x_i)$ and $\phi(x_j)$ cannot be explicitly expressed, the weight matrix W^ϕ in kernel space cannot be directly calculated. With the help of kernel function, the distance between $\phi(x_i)$ and $\phi(x_j)$ in kernel space can be transformed into function of vectors in the original space, i.e.

$$\begin{aligned} \left\|\phi(x_i)-\phi(x_j)\right\|^2 &= <\phi(x_i)-\phi(x_j), \phi(x_i)-\phi(x_j)> \\ &= <\phi(x_i),\phi(x_i)> - 2<\phi(x_i),\phi(x_j)> + <\phi(x_j),\phi(x_j)> \\ &= k(x_i,x_i) - 2k(x_i,x_j) + k(x_j,x_j) \end{aligned} \tag{8}$$

Thus (7) is transformed as follows:

$$W^{\phi}_{ij} = \begin{cases} \exp((-k(x_i,x_i)+2k(x_i,x_j)-k(x_j,x_j))/t) & \text{if } x_i \text{ and } x_j \text{ belong to the same class} \\ 0 & \text{otherwise} \end{cases} \quad (9)$$

It is easy to know that $P^{\phi} = \sum_{i=1}^{d} \alpha_i \phi(x_i) = X^{\phi}\boldsymbol{\alpha}$, where $\boldsymbol{\alpha} = [\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_d]$, d is the feature number.

Then, (6) can be converted into

$$\min_{P^{\phi}} \sum_{ij} \left\| y_i^{\phi} - y_j^{\phi} \right\|^2 W_{ij}^{\phi} = 2\boldsymbol{\alpha}^T X^{\phi T} X^{\phi} L^{\phi} X^{\phi T} X^{\phi} \boldsymbol{\alpha}$$
$$= 2\boldsymbol{\alpha}^T K L^{\phi} K \boldsymbol{\alpha} \quad (10)$$

where $K(x_i,x_j)$ is composite kernels as in (4), K_w and K_s are the spectral and spatial kernel matrices, respectively. The spatial information and spectral information from original hyperspectral image are combined using this composite kernel.

Note that, the matrix D^{ϕ} provides a natural measure on the data points. If D_{ii}^{ϕ} is large, then it implies that the class containing $\phi(x_i)$ has a high density around $\phi(x_i)$. Therefore, the bigger the value of D_{ii}^{ϕ} is, the more 'important' is $\phi(x_i)$. Therefore, we impose a constraint as follows:

$$y^{\phi T} D^{\phi} y^{\phi} = 1 \Rightarrow X^{\phi^T} P^{\phi} D^{\phi} P^{\phi T} X^{\phi} = 1$$
$$\Rightarrow \boldsymbol{\alpha}^T X^{\phi T} X^{\phi} D^{\phi} X^{\phi T} X^{\phi} \boldsymbol{\alpha} = 1 \Rightarrow \boldsymbol{\alpha}^T K D^{\phi} K \boldsymbol{\alpha} = 1 \quad (11)$$

Thus, the objective function (10) becomes the following:

$$\min_{P^{\phi}} \boldsymbol{\alpha}^T K L^{\phi} K \boldsymbol{\alpha} \quad (12)$$
$$\text{Subject to} \quad \boldsymbol{\alpha}^T K D^{\phi} K \boldsymbol{\alpha} = 1$$

This minimization problem can be converted to a generalized eigenvalue problem

$$K L^{\phi} K \boldsymbol{\alpha} = \lambda K D^{\phi} K \boldsymbol{\alpha} \quad (13)$$

Let the column vector $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_d$ be the solutions of equation (13), ordered according to their eigenvalues $\lambda_1 < \lambda_2 < \cdots \lambda_d$. Thus, the projection of a point $\phi(z)$ is as follows:

$$P^{\phi T} \phi(z) = \boldsymbol{\alpha}^T X^{\phi T} \phi(z) = \boldsymbol{\alpha}^T \begin{bmatrix} \phi(x_1)^T \\ \phi(x_2)^T \\ \vdots \\ \phi(x_m)^T \end{bmatrix} \phi(z) = \boldsymbol{\alpha}^T \begin{bmatrix} k(x_1,z) \\ k(x_2,z) \\ \vdots \\ k(x_m,z) \end{bmatrix} \quad (14)$$

Note that the kernel matrix K is an $m \times m$ matrix. Hence, if $m>d$, then K is a positive semidefinite matrix, i.e., K is singular. In this case, one may do decomposition (eigen-decomposition, QR decomposition, etc.) of K.

SCKLPP procedure is summarized in the following steps.

1) Compute the spectral and spatial kenerl matrix K_w, K_s, and get the kernel matrix K according to the given kernel function.

2) Compute the weight matrix W^{ϕ} and, hence, get L^{ϕ}, D^{ϕ}.

3) Extract features by solving $K L^{\phi} K \boldsymbol{\alpha} = \lambda K D^{\phi} K \boldsymbol{\alpha}$, and get $\boldsymbol{\alpha} = [\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_d]$.

4) Compute P^{ϕ} and the projections of sample points.

V. EXPERIMENTAL RESULTS

In this paper, a mixed forest/agricultural site in Indiana, is applied to compare the performances of SCKLPP with several feature extraction methods including PCA, LDA, and SKLPP with the only spectral information.

A. Data set

The AVIRIS Indian Pines image was taken over Northwest Indiana's Indian Pines test site in June 1992. The calibrated data are available online (along with the detailed ground-truth information) from http://dynamo.ecn.purdue.edu/~biehl/. As in [16], we first use a part of the scene, call the subset scene, consisting of pixels [31−116]×[27−94] for a size of 86×68, which contains four labelled classes (Fig. 1). Second, we use the whole scene, consisting of the full 145×145 pixels, which contain 16 classes. This study uses nine categories (Fig. 2): Corn-min (C1), Corn-notill (C2), Grass/Pasture (C3), Grass/Tree (C4), Hay-windrowed (C5), Soybeans-min (C6), Soybeans-clean (C7), Soybeans-notill (C8), and Woods (C9). These two scenes have unclassified background segments which don't have a representative spectra, and background segmentation is not addressed in our test. To overcome this, we force each pixel in the image to take a label of one of the existing 4/9 classes. The background pixels are excluded while considering the accuracy of the classification. In all images, we remove 35 noisy bands covering the region of water absorption, and finally work with 185 spectral bands. The AVIRIS 185-band image is first reduced by using the proposed SCKLPP method. Once the dimension reduced (projected) image is obtained, we perform the classification by using maximum likelihood (ML) classifier.

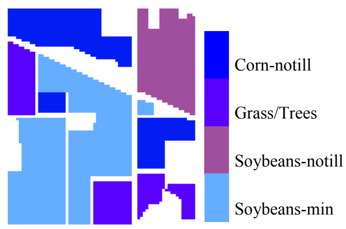


Figure 1 The ground truth data of subset image

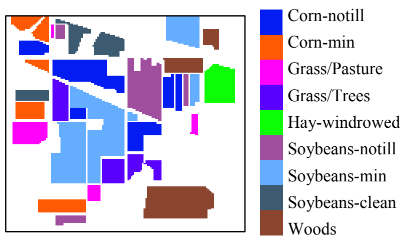


Figure 2 Ground truth of whole image with nine classes

B. Parameters Selection

For all the feature extraction method, the dimension of the subspace d is tuned within $d \in \{3,4,\cdots,35\}$. For SKLPP and SCKLPP, regularization parameter μ was varied in steps of 0.1 in the range [0,1], and the neighborhood size k between 1 and 15 is tested. In all cases, we used the polynomial kernel $d' \in \{1,2,\cdots,10\}$ for the spectral features and the RBF kernel $t \in \{10^{-1},\cdots,10^{3}\}$ for the spatial features. The spatial kernel is built using the mean of the neighborhood pixels in a 5×5 window $(\dim(x_i^s)=185)$ per spectral channel. In the following, we denote SKLPP_R as SKLPP with the RBF kernel.

C. Experimental results for the subset image

The number of training samples and testing samples in experiment 1 are presented in Table I. Table Π compares our approach with equivalent result from other feature methods with ten replications using random training samples. Note that the best accuracy of each data set is highlighted in shadow cell. The number in the parentheses is the number of features used for obtaining the best classification accuracy. Fig. 3 is thematic maps of using ML classifier with different feature extraction method which is the combination with highest classification accuracy.

TABLE I. NUMBER OF TRAINING SAMPLES AND TESTING SAMPLES USED IN EXPERIMENT 1

Class	Training	Test
C1 - Corn-notill	50	955
C2 - Grass/Trees	35	695
C3 - Soybeans-notill	50	682
C4 - Soybeans-min	90	1813
Total	**225**	**4145**

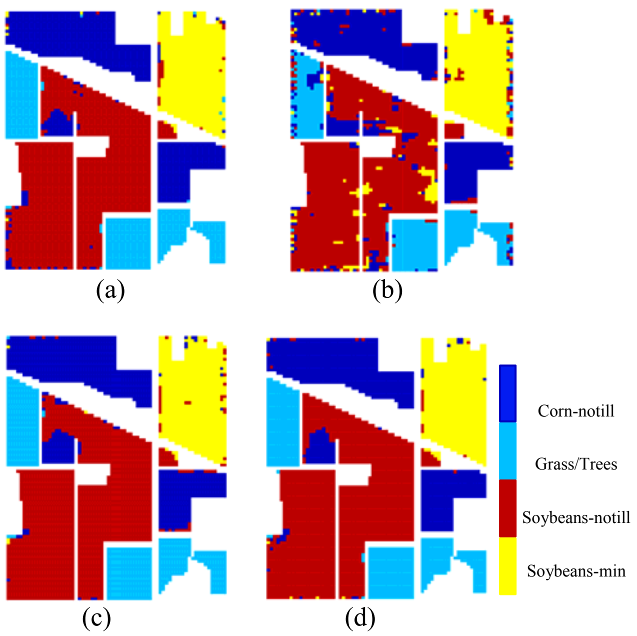


Figure 3 Thematic maps of using ML classifier with different feature extraction method in experiment 1, respectively. (a) PCA+ML. (b) LDA+ML. (c) SKLPP_R+ML. (d) SCKLPP+ML.

TABLE II. RESULTS FOR THE SUBSET IMAGE. PRODUCER'S ACCURACY (%), OVERALL ACCURACY (OA [%]), KAPPA STATISTIC (Κ) AND THE REDUCED FEATURE DIMENSIONS (D) IN TEST SET FOR DIFFERENT FEATURE EXTRACTION METHOD WITH ML CLASSIFIER.

Class	PCA	LDA	SKLPP_R	SCKLPP
C1	92.74	91.84	**97.31**	97.21
C2	**100.0**	88.63	**100.0**	**100.0**
C3	**96.31**	88.11	94.67	95.90
C4	95.74	87.18	97.16	**98.06**
OA	97.00	88.65	97.25	**97.88**
κ	0.9409	0.8390	0.9608	**0.9690**
d	13	3	15	13

From Fig. 3, it can be observed that the SCKLPP method reduced misclassifications and uncertainty in homogeneous regions. Form Table 2, we can find that our SCKLPP method produces better results than other feature extraction methods, especially in class C4. This shows that by considering the spatial information into hyperspectral subspace extraction process, it can be enhance the classification accuracy.

D. Experimental results for the whole image

In this experiment, 10% labelled samples of per class is used for training and the rest for validation. Fig. 4 is graphs of accuracies of using ML classifier with different feature extraction method in experiment, respectively. The thematic maps resulting from the classification of dimension reduced image using ML classifier which is the combination with highest classification accuracy are shown in Fig.5.

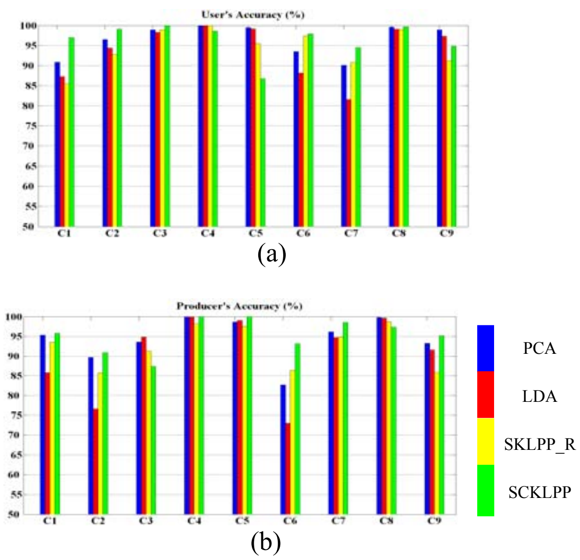
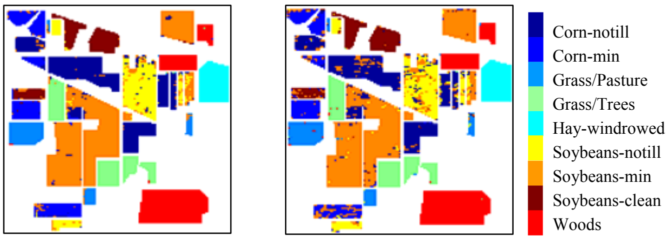
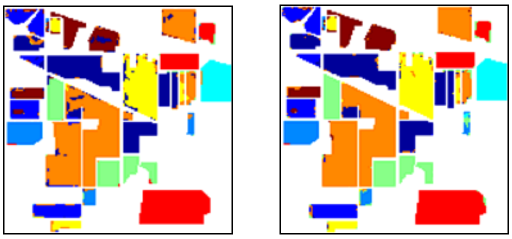


Figure 4 Graphs of accuracies of using ML classifier with different feature extraction method in experiment, respectively. (a) User's accuracies. (b) Producer's accuracies.

From Fig. 4, we can find that the user's accuracy of SCKLPP outperform other feature extraction methods on classes C1, C2, and C7, the user's accuracy improvements are 11.13%, 6.25%, 12.94% compared to SKLPP_R and LDA, while producer's accuracy on C1, C2, C6, C7, and C9, the producer's accuracy improvements are 9.95%, 14.21%, 20.16%, 3.86%, 9.27% compared to LDA and SKLPP_R, respectively.



(a) PCA (OA=94.54%, κ=0.9357, *d*=22) (b) LDA (OA=90.28%, κ=0.8850, *d*=8)



(c) SKLPP_R(OA=92.81%, κ=0.9154, *k*=17, *t*=680, *d*=28) (d) SCKLPP (OA=**95.82%**, κ=**0.9509**, *k*=19, *d'* =5, *t*=350, μ=0.7, *d*=22)

Figure 5 Thematic maps of using ML classifier with different feature extraction method in experiment 2, overall accuracies (OA [%]), Kappa statistic (κ), and the corresponding value of different parameters, respectively. (a) PCA+ML. (b) LDA+ML. (c) SKLPP_R +ML. (d) SCKLPP +ML.

From Fig. 5, we can see that SCKLPP has the best visual result, particularly in the area of Grass/trees, Soybeans-notill and Soybeans-min. By considering the spatial information into hyperspectral subspace extraction process, our SCKLPP outperform SKLPP algorithm, and the overall accuracies enhance 2.28%.

VI. CONCLUSIONS

To exploit both spectral and spatial information in feature extraction of hyperspectral remote sensing image, a kernel based manifold learning feature extraction method which considers spatial relationship of neighboring pixels, named SCKLPP, is proposed in this paper. The local geometric relations of the within-class samples in nonlinear kernel feature space are preserved. Experimental results on AVIRIS data sets show the efficiency of the proposed SCKLPP. The proposed method opens a new field for future developments in which spatial information can be easily integrated into the feature extraction stage.

ACKNOWLEDGMENT

We sincerely thank Prof. Landgrebe for providing the Indian Pine data set. Financial supports from the National Natural Science Foundation of China (61201323, 11202161) are gratefully acknowledged.

REFERENCES

[1] L. O. Jimenez, J. L. Rivera-median, E. Rodirguez-diaz, et al. "Integration of spatial and spectral information homgenous by means of unsupervised extraction and classification for objects applied to multispectral and hyperspectral data", IEEE Transaction on Geoscience and Remote Sensing, 43, pp. 844-851, 2005.

[2] C. M. Bachmann, T. L. Ainsworth and R. A. Fusina. "Exploiting manifold geometry in hyperspectral imagery", IEEE Transaction on Geoscience and Remote Sensing, 43, pp.441-454, 2005.

[3] T. Han, and G. Goodenough. "Nonlinear feature extraction of hyperspectral data based on Locally Linear Embedding (LLE)", IEEE Geoscience and Remote Sensing Symposium, pp. 1237-1240, 2005.

[4] Y. C. Chen, M. M. Crawford and J. Ghosh. "Applying nonlinear manifold learning to hyperspectral data for land cover classification", IEEE Geoscience and Remote Sensing Symposium, pp. 4311-4314, 2005.

[5] A. Mohan, G. Sapiro and E. Bosch. "Spatially coherent nonlinear dimensionality reduction and segmentation of hyperspectral images", IEEE Geoscience and Remote Sensing Letters, 4, pp. 206-210, 2007.

[6] G. Y. Chen, and S. E. Qian. "Dimensionality reduction of hyperspectral imagery using improved locally linear embedding", Journal of Applied Remote Sensing, 1, 2007.

[7] S. E. Qian, and G. Y. Chen. "A new nonlinear dimensionality reduction method with application to hyperspectral image analysis", IEEE International Geoscience and Remote Sensing Symposium, pp. 270-273, 2007.

[8] X. F. He and P. Niyogi. "Locality preserving projections", Advances in Neural Information Processing System 16, MIT Press, Cambridge, MA, pp. 153-160, 2003.

[9] J. Cheng, Q. Shan, H. Lu et al. "Supervised kernel locality preserving projections for face recognition", Neurocomputing, 67, pp. 443-44, 2005.

[10] X. Huang, L. Zhang. "A comparative study of spatial approaches for urban mapping using hyperspectral ROSIS images over Pavia city, northern of Italy", International Journal of Remote Sensing, 30, pp. 3205-3221, 2009.

[11] L. P. Zhang, X. Huang. "Object-oriented subspace analysis for airborne hyperspectral remote sensing imagery", Neurocomputing, 73, pp. 927-936, 2010.

[12] G. Camps-Valls, L. Gomez-chova. "Composite kernels for hyperspectral image classification", IEEE Transaction on Geoscience and Remote Sensing Letters, 3, pp. 93-97, 2006.

[13] V. F. Santiago, V. Manian. "Improving hyperspectral image classification using spatial preprocessing", IEEE Transaction on Geoscience and Remote Sensing Letters, 6, 297-301, 2009.

[14] M. Belkin and P. Niyogi. "Laplacian eigenmaps and spectral techniques for embedding and clustering", In Advances in Neural Information Processing System, 14, Cambridge, MA, USA, pp. 585-591, 2002.

[15] V. F. Santiago, J. Angulo. "Morphological scale-space for hyperspectral image and dimensionality exploration using tensor modeling", Hyperspectral Image and Signal Processing: Evolution in Remote Sensing, 2009, WHISPERS '09, 1-4.