UNSUPERVISED FEATURE EXTRACTION FOR HYPERSPECTRAL IMAGES USING COMBINED LOW RANK REPRESENTATION AND LOCALLY LINEAR EMBEDDING

Mengdi Wang, Jing Yu, *Member, IEEE*, Lijuan Niu, and Weidong Sun, *Member, IEEE*

***Abstract*—Hyperspectral images(HSIs) provide hundreds of narrow spectral bands for the land-covers, thus can provide more powerful discriminative information for the land-cover classification. However, HSIs suffer from the curse of high dimensionality, therefore dimension reduction and feature extraction are essential for the application of HSIs. In this paper, we propose an unsupervised feature extraction method for HSIs using combined low rank representation and locally linear embedding (LRR_LLE). The proposed method can simultaneously use both the spectral and spatial correlation within HSIs, with LRR modelling the intrinsic property of union of low-rank subspaces and LLE considering the correlation within spatial neighbours. Experiments are conducted on real HSI datasets and the classification results demonstrate that the features extracted by LRR_LLE are more discriminative than the state-of-art methods.**

***Index Terms*—Hyperspectral image, low rank representation, unsupervised feature extraction, locally linear embedding**

I. INTRODUCTION

Hyperspectral images (HSIs) can provide not only spatial but also spectral information of the land-covers in a scene and therefore are widely used in both civilian and military domain. However due to the existence of hundreds of spectral bands, HSIs always suffer from the curse of dimensionality [1]. According to [2], the increase in data dimensionality requires an exponential increase in the number of training samples when used for classification. However, there is always limited training data in HSIs. As a result, reducing the dimensionality while simultaneously remaining the structural discriminative features is rather important for HSIs.

In the past decades, different methods based on different constraints were proposed for the feature extraction and dimension reduction for HSIs. The linear algorithms based on principle component analysis (PCA) [3] and independent component analysis (ICA) [4] have been widely used due to their relative simplicity and effectiveness. PCA intends to preserve the most data variance with orthogonal projective subspaces, while ICA aims to find the statistical independent constitutional components using the mutual information as a criterion for the independence. Green *et. al* [5] proposed the maximum noise fraction (MNF) method with the measurement by signal-to-noise ratio (SNR), which can actually be seen as the noise-adjusted PCA. Nonlinear methods, such as ISOMAP [6], locally linear embedding (LLE) [7], Laplacian Eigenmap [8], were proposed to preserve the correlation between samples in the lower dimensional manifold. Several methods were proposed based on the mixing nature of HSIs, such as vertex component analysis (VCA) [9] and minimum volume constrained nonnegative matrix factorization (MVC-NMF) [10], both of which were firstly proposed for hyperspectral unmxing and as a result can serve as dimension reduction methods. The mostly recently work intrinsic representation (IR) [11] employed the underlying physical mixing factors and addressed the noise variance spectrally heterogeneity effect and the spatial correlation at the same time.

The methods above all aim to explore a representation that can simultaneously reduce the dimensionality and extract the structural discriminative features. In accordance with this basic principle, we propose a novel method using the framework of low rank representation (LRR), as LRR can structurally represent the spectral space in HSIs in an unsupervised way, which is a union of multiple low-rank subspaces. At the same time, LLE is combined with LRR to address the spatial correlation information.

The rest of this paper is organized as following. Section 2 introduces the novel feature extraction method LRR_LLE. Experimental results and discussion are shown in Section 3, and conclusion is drawn in Section 4.

M. Wang and W. Sun are with the State Key Laboratory of Intelligent Technology and Systems, Tsinghua National Laboratory for Information Science and Technology, Department of Electronic Engineering, Tsinghua University, Beijing 100084, China (e-mail: wangmd12@mails.tsinghua.edu.cn; wdsun@tsinghua.edu.cn).

J. Yu is with the College of Computer Science and Technology, Beijing University of Technology, Beijing 100124, China (e-mail: jing.yu@bjut.edu.cn).

L. Niu is with Cancer Hospital of Chinese Academy of Medical Sciences, Beijing 100021, China.

This work was partly supported by the National Nature Science Foundation (No.61501008) and the Key Project of the Science & Technology Development Program of BEC (No.kz201310028035) of China.

II. THE PROPOSED METHOD

A. The framework of LRR

An HSI data $\mathbf{X} \in \mathbb{R}^{P \times B}$ (which is reorganized from the HSI datacube $\mathcal{X} \in \mathbb{R}^{M \times N \times B}$ with $P = MN$ denoting the pixel number and B being the band number), is always considered lying in a low-dimensional manifold, which can be derived from the linear mixing model (LMM) [10], [12], [13]. The pixels in HSI can always be classified into several categories according to their spectral property, such as soil, vegetation, water, road, and buildings. Based on the class property, the spectral space of HSI can be divided into multiple subspaces, termed as $\{\mathcal{S}\}_c^K$. The spectral vectors within each class are highly correlated and thus they should lie in a low-dimensional manifold, which means that $\mathcal{S}_c$ is low-rank. Therefore the whole spectral space can be seen as the union of such low-rank subspaces $\bigcup_{c=1}^K \mathcal{S}_c$. Considering the mixing nature of hyperspectral pixels, the spectral data may lie in a union of multiple low-rank subspaces instead of a single low-rank space. An informative data representation when used for feature extraction should preserve the subspace-inherent structures and minimize the inter-subspace components. LRR [14] is a powerful representation tool for this purpose, because it is demonstrated that LRR is able to represent the structural information of the union of multiple independent subspaces [14]. Data drawn from $\bigcup_{c=1}^K \mathcal{S}_c$ can be handled with the following formulation:

$$\min_{\mathbf{Z},\mathbf{E}} \operatorname{rank}(\mathbf{Z}) + \lambda\|\mathbf{E}\|_0 \quad \text{s.t. } \mathbf{X} = \mathbf{A}\mathbf{Z} + \mathbf{E} \tag{1}$$

wherein, $\mathbf{A}$ is a dictionary that can span the data subspaces, $\|\cdot\|_0$ denotes the ℓ_0-norm of the noise matrix $\mathbf{E}$ and λ is a balance parameter. The structural reconstructed data can be obtained by $\hat{\mathbf{X}} = \mathbf{A}\mathbf{Z}^*$, where $\mathbf{Z}^*$ is the optimal solution of Eq.(1). However, in real application, the dictionary $\mathbf{A}$ is always unknown, a good choice is to set $\mathbf{A} = \mathbf{X}$. To make the optimization convex, the rank constraint and the ℓ_0-norm are replaced by nuclear norm and ℓ_1-norm respectively,

$$\min_{\mathbf{Z},\mathbf{E}} \|\mathbf{Z}\|_* + \lambda\|\mathbf{E}\|_1 \quad \text{s.t. } \mathbf{X} = \mathbf{X}\mathbf{Z} + \mathbf{E} \tag{2}$$

Eq.(2) can be solved using the inexact augmented Lagrange multiplier(IALM) [15] method.

B. Unsupervised feature extraction using LRR_LLE

The framework of LRR can only represent the global spectral structure of HSIs. However as pointed out in [16], it is essential to integrate the spatial correlation together with the spectral correlation when extracting features from HSIs, as the correlation within neighboring pixels can be potentially high.

LLE was proposed in [7] for unsupervised feature extraction based on simple geometric intuition that the data point and its neighbours are expected to lie on or close to a locally linear patch of a manifold. Therefore LLE is a good framework to address the spatial correlation in HSIs.

In this section, we will firstly outline the LLE method and secondly describe how we combine LRR and LLE together for the unsupervised feature extraction.

1) Locally linear embedding: LLE is proposed to compute low-dimensional, neighborhood-preserving embeddings of high-dimensional inputs. It can recover global nonlinear structure from locally linear fits. The local geometry is characterized by linear coefficients that reconstruct the data point using its neighbours. The coefficients are obtained by the following optimization,

$$\{W_{ij}\} = \arg\min_{W_{ij}} \|\mathbf{X}_i - \sum_j W_{ij}\mathbf{X}_j^{(i)}\|_{\mathrm{F}}^2 \tag{3}$$

where $\mathbf{X}_i$ is the data point (column-vector) to be reconstructed and $\{\mathbf{X}_j^{(i)}\}$ are the neighbours of $\mathbf{X}_i$. W_{ij} summarizes the contribution of $\mathbf{X}_j^{(i)}$ to the reconstruction of $\mathbf{X}_i$. According to LLE, the extracted features should preserve neighbourhood geometric manifold, therefore the embedding cost function is,

$$\mathrm{L} = \sum_i \|\mathbf{Y}_i - \sum_j W_{ij}\mathbf{Y}_j^{(i)}\|_F^2 = \operatorname{Tr}\left(\mathbf{Y}(\mathbf{I}-\mathbf{W})^{\mathrm{T}}(\mathbf{I}-\mathbf{W})\mathbf{Y}^{\mathrm{T}}\right) \tag{4}$$

where $\mathbf{Y}_i$ is the extracted feature, or somehow a transform in the new manifold, from data point $\mathbf{X}_i$, and $\mathbf{Y}_j^{(i)}$ corresponds to $\mathbf{X}_j^{(i)}$, $\mathbf{Y}$ is the matrix constituted by the set of vectors $\{\mathbf{Y}_j\}$, $\mathbf{W}$ is coefficient matrix, with $[\mathbf{W}]_{ij}$ being W_{ij} if $\mathbf{X}_j$ is a neighbour of $\mathbf{X}_i$ and 0 if not.

2) The LRR_LLE method: In the proposed method, we aim to combine the LRR and LLE for the feature extraction of HSIs, while LRR can employ the structure of the union of low-rank subspaces and LLE can take the spatially correlation into account. In the model described in Eq.(2), the ith-column $\mathbf{Z}_i$ in the coefficient matrix $\mathbf{Z}$ is actually a transform for the ith data sample $\mathbf{X}_i$ in the self-representation domain. Therefore, if $\mathbf{X}_j$ is a neighbour of $\mathbf{X}_i$ and contributes a weight with W_{ij} from model (3), $\mathbf{Z}_j$ should also be the neighbour of $\mathbf{Z}_i$ with the weight W_{ij}. From the framework of LLE, the embedding constraint should be $\operatorname{Tr}\left(\mathbf{Z}(\mathbf{I}-\mathbf{W})^{\mathrm{T}}(\mathbf{I}-\mathbf{W})\mathbf{Z}^{\mathrm{T}}\right)$. Therefore, the whole LRR_LLE method is modelled as,

$$\min_{\mathbf{Z},\mathbf{E}} \|\mathbf{Z}\|_* + \lambda\|\mathbf{E}\|_1 + \frac{\beta}{2}\operatorname{Tr}\left(\mathbf{Z}(\mathbf{I}-\mathbf{W})^{\mathrm{T}}(\mathbf{I}-\mathbf{W})\mathbf{Z}^{\mathrm{T}}\right) \quad \text{s.t. } \mathbf{X} = \mathbf{X}\mathbf{Z} + \mathbf{E} \tag{5}$$

where β is a balance parameter between the low-rank constraint and the LLE manifold constraint. The correlation matrix $\mathbf{W}$ can be constructed using the geometric reconstruction of $\mathbf{X}$. The neighbours of $\mathbf{X}_i$ are selected by the geometric distance measurement in the image domain, which means that the pixels within a grid size of $w \times w$ in the neighbourhood of $\mathbf{X}_i$ are selected. Eq.(5) can also be solved using IALM [15].

The structural extracted data can be obtained using $\hat{\mathbf{X}} = \mathbf{X}\mathbf{Z}^*$, where $\mathbf{Z}^*$ is the optimal solution of Eq.(5). However this is just a data processing procedure, with the data dimensionality remaining unchanged. For purpose of feature extraction and dimension reduction, PCA is used as a post-process procedure.

III. Experiments and discussion

In the experiments, the proposed LRR_LLE method is tested on two real HSI datasets. The evaluation of the effectiveness of the method is through HSI classification accuracies using supported vector machine (SVM) with the radial basis function (RBF).

A. Experiments set-up

The experiments set-up mainly contains two aspects: datasets and compared methods.

1) Datasets: There are two datasets used in the experiments: 1) The Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) image Indian Pines was taken over Northwest Indianas Indian Pines test site in June 1992. Indian Pines has a spatial size of 145×145 and 220 bands. Considering the water absorption, bands [104-108], [150-163], 220 are heavily corrupted and removed from the original data, therefore a $145 \times 145 \times 200$ data is used in the experiments. There are 16 classes in the data. A sample band is shown in Fig.1(a). 2) The data of Pavia University acquired by the reflective optics system imaging spectrometer(ROSIS) is used, with 103 bands and 610×340 pixels. Within Pavia University there are 9 classes of land-covers. Fig.1(b) shows the band 83 of the Pavia University data.

(a) (b)

Fig. 1. Sample bands of the datasets: (a) band 100 of the Indian Pines, (b) band 83 of the Pavia University.

2) Compared methods: Two traditional widely used methods, PCA and ICA; a hyperspectral unmixng-based method, MVC-NMF; and the most recently work IR [11] (only the complete method referred as IR2 in the paper are used here), which is the state-of-art method as long as we know, are used as the compared methods. To evaluate the effects of different parts of our proposed method, the results of LLE, LRR and LRR_LLE are also compared.

B. Experimental results and discussion

For the Indian Pines, each unsupervised feature extraction method is applied to extract 20 features. There are 16 classes in Indian Pines and 10% of the labelled samples in each class are used for training and the rest for testing. The classification are repeated for 10 times and the average of the results, including the average accuracy (AA), overall accuracy (OA) and the kappa coefficient, are shown in Table I. It can be observed that the proposed method LRR_LLE performs best, outperforming the state-of-art method IR for about 5.6% in OA and 5.2% in AA. Compared with the original data, the data after feature extraction and dimension reduction using LRR_LLE improves OA by 11.37% and AA by 12.54%. Both LLE and LRR performs worse than the original data, however the combined method LRR_LLE method performs much better. It means that LRR_LLE benefits from the integration of the spectrally constraint LRR and spatially constraint LLE.

All the methods are used for Pavia University to extract 15 features, which serve as the input of the SVM classifier. There are 9 classes in Pavia University and 100 samples in each class are used as the training set and the remaining as the testing set. It is shown that LRR, IR and LRR_LLE all outperforms the original data, while LRR_LLE is the best. Compared with IR, the proposed method achieves an improvement of about 1.9% and 1.1% respectively in OA and AA. LLE performs slightly worse than the original data and LRR performs better, however not good enough, even worse than IR. Compared with LLE and LRR, the great performance of LRR_LLE is demonstrated to be derived from the combination of the spectral and spatial constraint.

TABLE I

CLASSIFICATION ACCURACIES USING SVM (OA, AA IN PERCENTAGE) FOR THE ORIGINAL DATA AND THE RESULTS BY THE UNSUPERVISED FEATURE EXTRACTION METHODS (BEST RESULTS ARE HIGHLIGHTED IN BOLD).

	Indian Pines			Pavia University		
	OA	AA	kappa	OA	AA	kappa
original	82.76	80.76	0.8034	88.31	90.45	0.8479
PCA	79.95	79.87	0.7712	74.47	82.35	0.6777
ICA	74.27	70.71	0.7057	83.27	87.38	0.7840
MVC-NMF	74.04	71.12	0.7023	82.96	85.78	0.7775
LLE	79.76	77.47	0.7694	87.77	90.04	0.8411
LRR	82.34	78.47	0.7984	91.22	92.34	0.8852
IR	88.5	88.1	0.869	93.1	94.3	0.909
LRR_LLE	**94.13**	**93.30**	**0.9330**	**95.03**	**95.43**	**0.9345**

The impact of the number of features on the classification accuracy is analysed using the Indian Pines dataset. Fig.2 shows the OA and AA upon different number of the extracted features. As IR [11] does not provide the source code of the method, we only use the results in the reference paper using a dashed line, which is actually the result when the number of features is 20. The classification result of the original data can provide a reference accuracy, however it has no correlation with the number of features, therefore is also shown in a dashed line. It can be observed that the proposed method LRR_LLE greatly outperforms the other methods. It achieves the highest accuracy when the number of features is approximately 20 and remains approximately constant even when the number of features increases. LRR can outperform the original data when the number of features is comparatively large. PCA performs slightly worse than the original data and performs better when the number of feature increases. Apparently it has the upper bound by the result of the original data. LLE, ICA and MVC-NMF are all worse than the original data, while ICA and MVC-NMF has the best result when the number is approximately between 20-30. In conclusion, our proposed method has great performance and shows great stability upon the number of extracted features.

Fig.3 shows the classification accuracies of Indian Pines using different numbers of training samples per class, when the number of features are fixed at 20. It is shown that when the training samples are only 10 per class, our proposed method LRR_LLE performs slightly worse than IR, however in average LRR_LLE performs the best among all the methods.

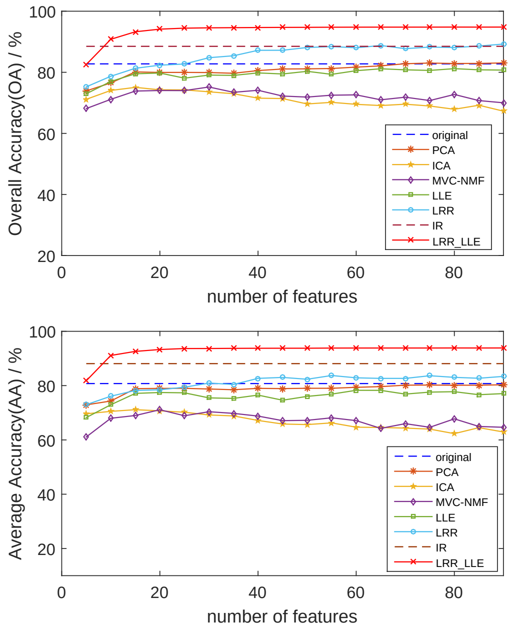

Fig. 2. Classification results for Indian Pines using different numbers of features.

LRR_LLE achieves an average improvement of 4.47% in OA and 2.97% in AA upon IR, which are 12.07% and 7.35% respectively when compared with the original data.

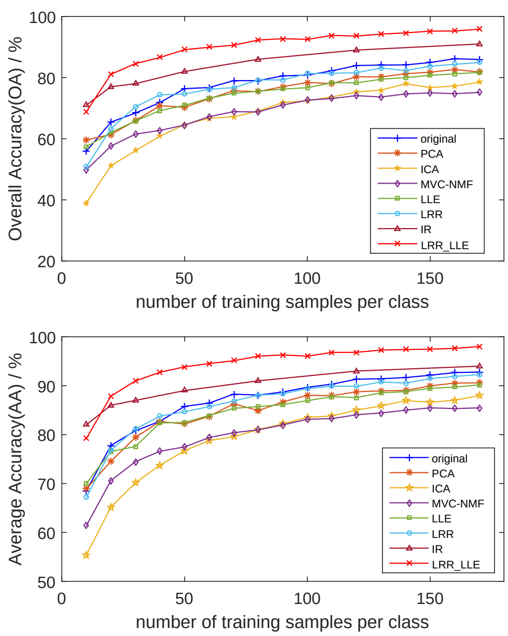

Fig. 3. Classification results for Indian Pines using different numbers of training samples per class.

IV. CONCLUSION

In this paper, we have proposed a novel unsupervised feature extraction method using combined LRR and LLE for HSIs. LRR is a framework for robust recovery of a union of multiple low-rank subspaces, therefore is capable to structurally represent the spectral space of HSIs. LLE is a nonlinear dimension reduction method which aims to preserve the locally geometric manifold. Combination of LRR and LLE using model (5) can simultaneously employ the structurally spectral correlation and the locally spatial correlation information. Experiments with a following classification task using SVM show that the proposed method LRR_LLE benefits from the combination of the spectral constraint using LRR and spatial constraint using LLE, and greatly outperforms the state-of-art methods when used for the structural unsupervised feature extraction for HSIs.

REFERENCES

[1] D. L. Donoho *et al.*, "High-dimensional data analysis: The curses and blessings of dimensionality," *AMS Math Challenges Lecture*, pp. 1–32, 2000.

[2] G. Hughes, "On the mean accuracy of statistical pattern recognizers," *IEEE Transactions on Information Theory*, vol. 14, no. 1, pp. 55–63, 1968.

[3] H. Hotelling, "Analysis of a complex of statistical variables into principal components." *Journal of Educational Psychology*, vol. 24, no. 6, p. 417, 1933.

[4] J. Wang and C.-I. Chang, "Independent component analysis-based dimensionality reduction with applications in hyperspectral image analysis," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 6, pp. 1586–1600, 2006.

[5] A. A. Green, M. Berman, P. Switzer, and M. D. Craig, "A transformation for ordering multispectral data in terms of image quality with implications for noise removal," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 26, no. 1, pp. 65–74, Jan 1988.

[6] C. M. Bachmann, T. L. Ainsworth, and R. A. Fusina, "Exploiting manifold geometry in hyperspectral imagery," *IEEE transactions on Geoscience and Remote Sensing*, vol. 43, no. 3, pp. 441–454, 2005.

[7] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.

[8] M. Belkin and P. Niyogi, "Laplacian eigenmaps and spectral techniques for embedding and clustering," in *Advances in Neural Information Processing Systems*, 2001, pp. 585–591.

[9] J. M. Nascimento and J. M. Dias, "Vertex component analysis: A fast algorithm to unmix hyperspectral data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 4, pp. 898–910, 2005.

[10] L. Miao and H. Qi, "Endmember extraction from highly mixed data using minimum volume constrained nonnegative matrix factorization," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 3, pp. 765–777, 2007.

[11] L. Xu, A. Wong, F. Li, and D. A. Clausi, "Intrinsic representation of hyperspectral imagery for unsupervised feature extraction," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 2, pp. 1118–1130, 2016.

[12] H. Zhang, W. He, L. Zhang, H. Shen, and Q.Yuan, "Hyperspectral image restoration using low-rank matrix recovery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 8, pp. 4729–4743, Aug 2014.

[13] M. Wang, J. Yu, J. H. Xue, and W. Sun, "Denoising of hyperspectral images using group low-rank representation," *IEEE J. Sel. Topics Appl. Earth Observ. in Remote Sens.*, vol. 9, no. 9, pp. 4420–4427, Sept 2016.

[14] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma, "Robust recovery of subspace structures by low-rank representation," *IEEE Transaction on Pattern Analysis Machine Intelligence*, vol. 35, no. 1, pp. 171–184, Jan 2013.

[15] Z. Lin, M. Chen, and Y. Ma, "The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices," in *arXiv preprint arXiv:1009.5055*, 2010.

[16] A. Plaza, J. A. Benediktsson, J. W. Boardman, J. Brazile, L. Bruzzone, G. Camps-Valls, J. Chanussot, M. Fauvel, P. Gamba, A. Gualtieri *et al.*, "Recent advances in techniques for hyperspectral image processing," *Remote Sensing of Environment*, vol. 113, pp. S110–S122, 2009.