SPECTRAL-SPATIAL HYPERSPECTRAL IMAGE CLASSIFICATION VIA LOCALITY AND STRUCTURE CONSTRAINED LOW-RANK REPRESENTATION

Xiang He[1], Qi Wang,[1,2], Xuelong Li[3]*

[1]School of Computer Science and Center for OPTical IMagery Analysis and Learning, Northwestern Polytechnical University, Xi'an 710072, Shaanxi, P. R. China.

[2]Unmanned System Research Institute, Northwestern Polytechnical University, Xi'an 710072, Shaanxi, P. R. China.

[3]Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an 710119, Shaanxi, P. R. China, and University of Chinese Academy of Sciences, Beijing 100049, P. R. China.

ABSTRACT

Low-rank representation (LRR) has been applied widely in most fields due to its considerable ability to explore the low-dimensional subspace embedding in high-dimensional data. However, there are still some problems that LRR can't effectively exploit the local structure and the representation for the given data is not discriminative enough. To tackle the above issues, we propose a novel locality and structure constrained low-rank representation (LSLRR) for hyperspectral image (HSI) classification. First, a distance metrics, which combines spectral and spatial similarity, is proposed to constrain the local structure. This makes two pixels in HSI with small distance have high similarity. Second, we exploit the classwise block-diagonal structure for the training data to learn the more discriminative representation for the testing data. And the experimental results verify the effectiveness and superiority of LSLRR comparing with other state-of-the-art methods.

Index Terms—hyperspectral image classification, low-rank representation, block-diagonal structure

1. INTRODUCTION

Hyperspectral images (HSIs) are acquired by imaging spectrometer, which contain abundant bands and cover a large range from visible to infrared spectrum. Compared with RGB image, HSI includes richer spectral information, which make HSI be applied widely in real applications. Among these, more attention have been drawn in hyperspectral image classification, which aims to assign each pixel in HSI an accurate class label according to training data.

In the past many years, many HSI classification methods have been developed, such as linear discriminate analysis (LDA) [1], support vector machine (SVM), multinomial logistic regression (MLR), and so on. However, these approaches only utilize the spectral feature to classify HSI, which strongly restricts the classification accuracy. Latterly some state-of-the-art methods combining both spectral and spatial information are presented and achieve a satisfying performance, such as locality adaptive discriminant analysis (LADA) [2], E_SVM [3], generalized composite kernel (GCK) [4].

Recently, Low-rank representation (LRR) has been initially proposed in [5] for subspace segmentation. Afterwards it's used successfully in various fields, e.g. face recognition [6], HSI classification [7]. Given a set of data, LRR can greatly capture the global structure by seeking the low-rank representation and the sparse noise based on some dictionary vectors. However, there are still two main drawbacks for HSI classification by LRR. First, LRR neglects the local data structure [8], which fails to preserve the neighborhood relations of pixels. Second, assume that all observed data lie in the union of several independent subspaces, then the optimal low-rank representation has an ideal block-diagonal structure. However, the assumption isn't always ensured, especially for HSI data.

To address the above mentioned problems, we propose a novel locality and structure constrained low-rank representation (LSLRR). Firstly, we introduce an integrated similarity metrics as the locality constraint which combines the spectral and spatial features. Secondly, we present a structure constraint to learn the discriminative representation for testing data by making the representation for training data be a classwise block-diagonal matrix.

2. LOCALITY AND STRUCTURE CONSTRAINED LOW-RANK REPRESENTATION (LSLRR)

2.1. Low-rank representation

Low-rank representation (LRR) [5] can effectively exploit the underlying low-dimensional subspace structures in original data. Consider data samples $Y \in \mathbb{R}^{d\times n}$, LRR aims to acquire the low-rank representation $Z \in \mathbb{R}^{m\times n}$ of Y based on the given dictionary $A \in \mathbb{R}^{d\times m}$. Specifically, LRR is formulated as follows

$$\min_{Z,E} \ rank(Z)+\lambda\|E\|_0 \quad s.t. \ Y=AZ+E, \qquad (1)$$

where A is the dictionary matrix, E is the noise component, which is generally sparse, $\|\cdot\|_0$ is the ℓ_0 norm, λ is the regularization coefficient. However, the above problem is non-convex because of the discrete nature of the rank operation and ℓ_0 norm. The relaxed problem of (1) is

$$\min_{Z,E} \ \|Z\|_* + \lambda \|E\|_{2,1} \quad s.t. \ Y = AZ + E, \tag{2}$$

where $\|\cdot\|_*$ is the nuclear norm, which is the sum of all singular values of Z, $\|\cdot\|_{2,1}$ is the $\ell_{2,1}$ norm, which is defined as $\|E\|_{2,1} = \sum_j^n \sqrt{\sum_i^d E_{i,j}^2}$.

2.2. Proposed LSLRR model

2.2.1. locality constraint for LSLRR

For HSI classification, two pixels with a neighboring relationship in similar regions have high possibility of belonging to the same class. That is, spatial similarity is a useful priori information to obtain more accurate results. What's more, LRR can effectively capture the global structure of given data, but the local geometric structure, also beneficial for classification, is ignored by LRR. Therefore, we utilize the spectral and spatial similarity to constrain the local structure.

Consider a HSI data $X = [x_1, x_2, ..., x_n] \in \mathbb{R}^{d \times n}$, where *d* is the number of bands and *n* is the number of all pixels. x_i denotes the spectral vector of the *i*-th pixel. Similarly, denoting that $L = [l_1, l_2, ..., l_n] \in \mathbb{R}^{2 \times n}$, and l_i denotes the spatial coordinate of *i*-th pixel. The way to exploit both spectral and spatial features is formulated as

$$D_{ij} = \sqrt{\|x_i - x_j\|_2^2 + m\|l_i - l_j\|_2^2}, \tag{3}$$

where *m* is a hyper-parameter to control the weight of spectral and spatial similarity. For different HSI data, the compactness of each class is different. For the HSI data with high compactness of each class, it's appropriate to set a large value for *m*. As we know, two pixels with a smaller distance should have a larger similarity. As such, the local constraint as a regularization term is proposed to preserve the compactness within classes and the separation between classes

$$\sum_{i,j} D_{ij} |Z_{ij}| = \|D \odot Z\|_1, \tag{4}$$

where ⊙ is the Hadamard product, denoting element-wise product, and $\|\cdot\|_1$ is the ℓ_1 norm.

2.2.2. structure constraint for LSLRR

Consider the data $X = [\bar{X}, \hat{X}] \in \mathbb{R}^{d \times n}$, where $\bar{X}$ is the training set and $\hat{X}$ is the testing set. More specifically, $\bar{X} = [X_1, X_2, ... X_c] \in \mathbb{R}^{d \times m}$ (*c* is the number of classes) is training feature matrix, and its training samples are rearranged according to each classes, i.e., X_i is the training data of the *i*-th class. In LRR model, we set the data $Y = [\bar{X}, \hat{X}]$, $A = \bar{X}$., so Z can be written as $[\bar{Z}, \hat{Z}]$.

Generally, the training set $\bar{X}$ can be reconstructed by the presentation $\bar{Z}$ based on dictionary $\bar{X}$ with a sparse noise. Because these data points $\bar{X}$ belonging to the same class are ordered, the ideal $\bar{Z}$ would be a classwise block-diagonal matrix.

To preserve the above structure for $\bar{Z}$, we introduce a structure constraint to make the non-diagonal blocks of $\bar{Z}$ are all zero sub-matrix. The detailed form of Q can be see the following section 2.3.

After combining the above locality and structure constraint terms with LRR, our proposed LSLRR can be formulated as

$$\begin{aligned} \min_{Z,E} \ & \|Z\|_* + \lambda \|E\|_{2,1} + \alpha \|D \odot Z\|_1 + \beta \|Z - Q\|_F^2 \\ \text{s.t.} \ & X = \bar{X}Z + E \end{aligned} \tag{5}$$

where λ, α, β are the regularization parameters.

2.3. ALM for solving LSLRR

Up to now, most methods [9, 10] have been developed to solve the LRR problem. Here we select the Augmented Lagrange Multiplier (ALM) for solving LSLRR. To make the problem (5) become separable, we first introduce two auxiliary variables J and H. Thus, (5) is converted to

$$\begin{aligned} \min_{H,J,Z,E} \ & \|J\|_* + \lambda \|E\|_{2,1} + \alpha \|D \odot H\|_1 + \beta \|Z - Q\|_F^2 \\ \text{s.t.} \ & X = \bar{X}Z + E, \ J = Z, \ H = Z \end{aligned} \tag{6}$$

Then the augmented Lagrangian function for (6) is written as follows

$$\begin{aligned} \min_{H,J,Z,E} \ & \|J\|_* + \lambda \|E\|_{2,1} + \alpha \|D \odot H\|_1 + \beta \|Z - Q\|_F^2 \\ & + \langle Y_1, X - \bar{X}Z - E \rangle + \langle Y_2, Z - J \rangle + \langle Y_3, H - Z \rangle \\ & + \frac{\mu}{2}(\|X - \bar{X}Z - E\|_F^2 + \|Z - J\|_F^2 + \|H - Z\|_F^2) \end{aligned} \tag{7}$$

where $\langle A, B \rangle = trace(A^T B)$, $\mu > 0$ is a penalty parameter, Y_1 and Y_2 are Lagrange multipliers. The problem (7) can be solved by alternative optimization approach.

Update H: fix the J, Z, E, then H can be updated as follows

$$H^{k+1} = \arg\min_H \frac{\alpha}{\mu^k} \|D \odot H^k\|_1 + \frac{1}{2} \|H^k - Z^k + \frac{Y_3^k}{\mu^k}\|_F^2 \tag{8}$$

The solution for (8) can be computed by

$$H_{ij}^{k+1} = \Theta_{w_{ij}}(Z_{ij}^k - \frac{Y_{3,ij}^k}{\mu^k}) \tag{9}$$

where $\Theta_w(x)=\max(x-w,0)+\min(x+w,0)$, $\omega_{ij}=(\alpha/\mu^k)D_{ij}$.

Update J: fix the H, Z, E, then J can be updated as follows

$$J^{k+1}=\arg\min_J \frac{1}{\mu^k}\|J^k\|_*+\frac{1}{2}\|Z^k-J^k+\frac{Y_2^k}{\mu^k}\|_F^2, \quad (10)$$
$$=US_{1/\mu^k}(\Sigma)V^T$$

where $U\Sigma V^T$ is the singular value decomposition (SVD) of $Z^k+Y_2^k/\mu^k$, and $S_\varepsilon(x)=\mathrm{sgn}(x)\max(|x|-\varepsilon,0)$ is the soft-thresholding operator.

Update Z: fix the H, J, E, then Z can be updated as follows

Denote $A=X-E+Y_1/\mu$, $B=J-Y_2/\mu$, $C=H+Y_3/\mu$, and $W=\mu I+2\beta I+\mu(\bar{X}^T\bar{X}+I)$. The optimal solution of Z is

$$Z^{k+1}=[W^k]^{-1}[2\beta Q^k+\mu^k(\bar{X}^TA^k+B^k+C^k)]. \quad (11)$$

And Q is updated by $Q^{k+1}=[\bar{Q}^k,\hat{Z}^k]$, where $\bar{Q}^k$ is defined as

$$\bar{Q}^k=\begin{bmatrix}\bar{Z}_1^k & 0 & 0 & 0\\ 0 & \bar{Z}_2^k & 0 & 0\\ 0 & 0 & \ddots & 0\\ 0 & 0 & 0 & \bar{Z}_c^k\end{bmatrix}, \quad (12)$$

Update E: fix the H, J, Z, then E can be updated as follows

$$E^{k+1}=\arg\min_H \frac{\lambda}{\mu^k}\|E^k\|_{2,1}+\frac{1}{2}\|X-\bar{X}Z^k-E^k+\frac{Y_1^k}{\mu^k}\|_F^2. \quad (13)$$

Denote $G=X-\bar{X}Z+Y_1/\mu$, the i-th column of optimal solution E is

$$E^{k+1}(:,i)=\begin{cases}\dfrac{\|g_i\|_2-\frac{\lambda}{\mu^k}}{\|g_i\|_2}g_i, & if\ \dfrac{\lambda}{\mu^k}<\|g_i\|_2\\ 0, & otherwise\end{cases}. \quad (14)$$

Finally, the complete algorithm for solving (5) is summarized as Algorithm 1.

Algorithm 1 ALM for solving LSLRR

Input: testing set X, training set $\bar{X}$, the local constraint matrix D, the structure constraint Q, parameter λ, α, β, m.

Initialize: H=J=Z=Q=0, E=0, $\mu=10^{-6}$, $max_\mu=10^{10}$, $\rho=1.1$, $\varepsilon=10^{-4}$, $Y_1=Y_2=Y_3=0$.

While not converged **do**

1. Compute optimal H, J, Z, E according to (9,10,11,14)
2. Update the multipliers
   $Y_1^{k+1}=Y_1^k+\mu k(X-\bar{X}Z^k-E^k)$
   $Y_2^{k+1}=Y_2^k+\mu^k(Z^k-J^k)$
   $Y_3^{k+1}=Y_3^k+\mu^k(H^k-Z^k)$
3. Update the parameter μ
   $u^{k+1}=\rho u^k$
4. Update the matrix Q by $Q^{k+1}=[\bar{Q}^k,\hat{Z}^k]$
5. Check the convergence conditions
   $\|X-\bar{X}Z-E\|_\infty<\varepsilon,\ \|Z-J\|_\infty<\varepsilon,\ \|H-Z\|_\infty<\varepsilon$
6. $k\leftarrow k+1$

End while

2.4. HSI classification via LSLRR

After solving the problem (5) by Algorithm 1, $\bar{Z}$ should be a classwise block-diagonal matrix, and the i-th column of $\hat{Z}$ is the new representation of the i-th testing sample based on all training data. Moreover, $\hat{z}_{ij}$ of coefficient matrix $\hat{Z}$ reflects the similarity of training sample x_i and testing sample x_j, so Z is sometimes considered an affinity matrix. For a testing pixel x_j, its label can be determined as follows. Firstly, for j-th testing pixel, compute the sum of j-th column of $\hat{z}$ for each class, denoted by $S_l(\hat{z}_j)$, $l\in[1,\ldots,c]$. Secondly, the label of x_j, denoted by *label(x_j)*, is confirmed as $label(x_j)=\arg\max_{l=1,\ldots,c} S_l(\hat{z}_j)$.

3. EXPERIMENTS

3.1. Dataset description and experimental setup

To verify the performance of the proposed LSLRR, we conduct the experiments on the popular Indian Pines dataset. It contains 16 classes and most of them are the agricultural regions. After removing some bands covering noise and water-absorption fields, the remaining number of spectral bands is 200.

The proposed LSLRR is compared with SVM, E_SVM [3], LRR [5], and LADA [2]. The parameters c and γ for SVM (using the RBF kernel) are determined by grid search algorithm. The regularization parameter λ for LRR is set as 0.35. And the parameters for the proposed LSLRR are $\lambda=0.1,\alpha=0.6,\beta=0.4,m=12$. We randomly select 10% of Indian Pines data for training and remaining 90% for testing. The overall accuracy (OA), average accuracy (AA) and kappa coefficient (κ) are adopted to evaluate the performance.

3.2. Experimental results and analysis

Because of the space limitation, a part of experimental results are shown in Table 1 and Fig.1. As shown in Table 1 (produces accuracy for each class), LSLRR can yield the better classification accuracy for most classes compared to other comparison methods, and all the three evaluation index for LSLRR is the highest. Besides, the classification map of LSLRR also obviously shows that only a small number of samples are classified wrongly. LADA obtains a satisfying

performance due to it can exploit the local manifold structure of data. And the proposed LSLRR performs so well. The main reasons are that the two powerful locality and structure constraints are integrated to LRR. This makes LSLRR be able to not only exploit the local structure, but also learn the discriminative representation for HSI data.

Table 1. CLASSIFICATION ACCURACY (%) FOR THE INDIAN PINES

Class	SVM	E_SVM	LRR	LADA	LSLRR
2	70.36	88.86	69.15	88.70	**91.96**
3	72.15	**94.62**	54.24	86.68	92.68
5	90.65	88.96	94.65	94.12	**95.23**
7	92.95	68.12	45.63	60.14	**98.76**
8	99.12	95.75	93.23	**100**	**100**
11	80.35	97.23	56.75	94.89	**97.43**
12	69.54	86.75	44.21	87.26	**93.15**
13	**96.25**	93.26	88.26	93.25	94.53
15	52.11	76.85	76.86	90.42	**94.26**
OA	80.49	90.56	64.26	91.63	**94.98**
AA	78.73	84.26	66.12	84.86	**94.15**
kappa	77.75	89.67	57.63	90.52	**93.95**

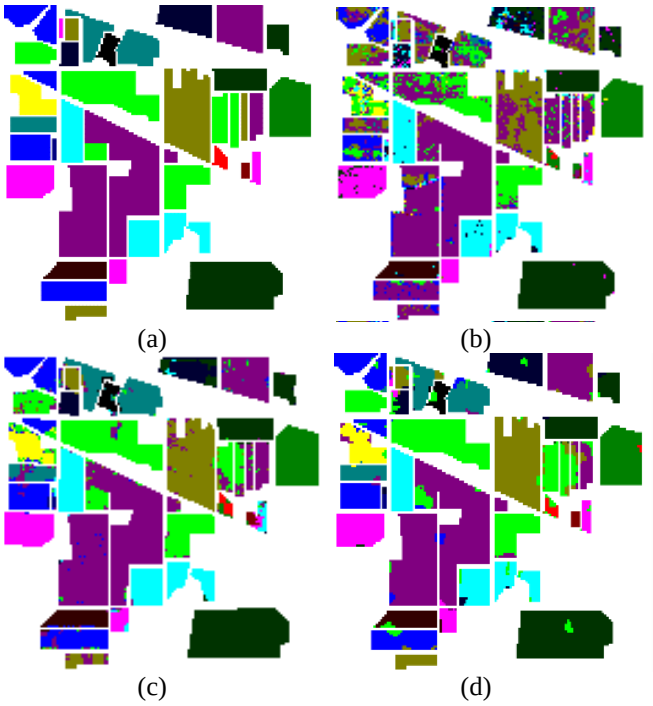

Fig. 1. Ground truth and classification maps for Indian Pines. (a) Ground truth, (b) LRR (64.26%), (c) LADA (91.63%), (d) LSLRR (94.98%).

4. CONCLUSION

In this paper, we propose a novel locality and structure constrained low-rank representation (LSLRR) for hyperspectral image classification. LSLRR with the two powerful constraints make it more suitable for HSI classification. And experimental results also demonstrate the effectiveness and superiority of the proposed LSLRR.

5. ACKNOWLEDGEMENT

This work was supported by the National Key R&D Program of China under Grant 2017YFB1002202, National Natural Science Foundation of China under Grant 61773316, Fundamental Research Funds for the Central Universities under Grant 3102017AX010, and the Open Research Fund of Key Laboratory of Spectral Imaging Technology, Chinese Academy of Sciences.

6. REFERENCES

[1] T. V. Bandos, L. Bruzzone, and G. Camps-Vall, "Classification of hyperspectral images with regularized linear discriminant analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 3, pp. 862–873, 2009.

[2] Q. Wang, Z. Meng, and X. Li, "Locality Adaptive Discriminant Analysis for Spectral–Spatial Classification of Hyperspectral Images," *IEEE Geosci. Remote Sens. Lett*, vol. 14, no. 14, pp. 2077-2081, 2017.

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

[4] J. Li, P. R. Marpu, and A. Plaza. "Generalized composite kernel framework for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 9, pp. 4816-4829, 2013.

[5] G. Liu, Z. Lin, and Y. Yu, "Robust subspace segmentation by low-rank representation," in *Proc. Int. Conf. Mach. Learn.*, 2010, pp. 663–670.

[6] Y. Li, J. Liu, Z. Li, et al. "Learning low-rank representations with classwise block-diagonal structure for robust face recognition," in *Proc. AAAI*, 2014, pp. 2810-2816.

[7] L. Pan, H. C. Li, and X. D. Chen, "Locality constrained low-rank representation for hyperspectral image classification," in *Proc. IGARSS*, 2016, pp. 493-496.

[8] Q. Wang, J. Wan, and Y. Yuan, "Locality Constraint Distance Metric Learning for Traffic Congestion Detection," *Pattern Recognition*, vol. 75, pp. 272-281, 2018.

[9] Z. C. Lin, M. M. Chen, and Y. Ma, "The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices," in *Technical report, UIUC Technical Report UILU-ENG-09-2215*, 2009.

[10] Z. C. Lin, R. S. Liu, and Z. X. Su, "Linearized alternating direction method with adaptive penalty for low-rank representation," in *Proc. NIPS*, Granda, ES, 2011, pp. 612–620.