THE K-LLE ALGORITHM FOR NONLINEAR DIMENSIONALITY RUDUCTION OF LARGE-SCALE HYPERSPECTRAL DATA

Danfeng Hong [a,b], *Naoto Yokoya* [a,b,c], *Xiao Xiang Zhu* [a,b]

[a] Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), Germany,
[b] Signal Processing in Earth Observation (SiPEO), Technical University of Munich, Germany
[c] Research Center for Advanced Science and Technology, The University of Tokyo, Japan

ABSTRACT

This work addresses nonlinear dimensionality reduction by means of locally linear embedding (LLE) for large-scale hyperspectral data. The LLE algorithm depends on spectral decomposition to a great extent, resulting in computational complexity and storage-costing while calculating the embedding of the low-dimensional data, particularly for large-scale hyperspectral data. LLE is not applicable to dimensionality reduction of large-scale hyperspectral data using general personal computers. In this paper, we present a novel method named K-LLE which introduces K-means clustering into LLE to deal with this issue. We firstly utilize K-cluster centers to represent the manifold structure of data instead of all data points, and next regard the K-cluster centers as a bridge between the manifold structure and all data in order to obtain the low-dimensional representation for each data point without handling the complex spectral decomposition. Finally, classification is explored as a potential application to validate the proposed algorithm. Experimental results on two hyperspectral datasets demonstrate the effectiveness and superiority of the proposed algorithm.

Index Terms— hyperspectral dimensionality reduction, large-scale, manifold learning, K-means clustering

1. INTRODUCTION

Recently, thanks to its very rich spectral information, hyperspectral imaging has been exploited for many remote sensing tasks, e.g., classification, detection. Rich spectral information provides the discriminative features; however, in the meantime brings the information redundancy. Therefore, as a common but necessary step, dimensionality reduction (DR) plays an important role in the hyperspectral data analysis.

Generally, the DR methods can be mainly categorized into linear and nonlinear methods. Compared to the linear DR techniques such as principle component analysis (PCA), manifold learning is a typically nonlinear DR approach. It tries to extract an intrinsic low-dimensional manifold structure from high-dimensional data by means of unsupervised strategies. Manifold learning is capable of embedding nonconvex and nonlinear manifolds in hyperspectral data by preserving the manifold structure from high-dimensional data to low-dimensional data. In particular, local manifold learning, e.g., locally linear embedding (LLE), achieves nonlinear DR through preserving the local neighborhood relationship. Extensive research achievements based on LLE have been reported in the field of hyperspectral image processing [1-3], which shows its superiority and effectiveness. However, DR based LLE is limited by LLE itself, i.e., LLE hardly copes with large-scale hyperspectral data [4]. Since the computational complexity and storage-costing of spectral decomposition is decisively related with the scale of hyperspectral data, general personal computers (PCs) are not able to directly process such large spectral decomposition.

This work aims at a general method to avoid calculating large-scale spectral decomposition, and in the meantime, embedding this idea into the LLE algorithm, enabling the algorithm to learn the underlying manifold structure from large-scale hyperspectral data. The novelty of our proposed K-LLE algorithm is: K cluster centers are used to represent the integral manifold of the original data, which means that the manifold structure can be learned only using K cluster centers rather than using all data points. More specifically, while computing the embedding of LLE algorithm, we preserve the local manifold structure between each point and K cluster centers in the low-dimensional space instead of considering local manifold structures of all points as in the original LLE algorithm. This leads to a significantly reduced demand on computational and storage-cost.

2. METHODOLOGY

2.1 The LLE Algorithm

The proposed method is based on the LLE algorithm. LLE can be briefly explained as follows. Let *D*-dimensional data $\mathbf{X}=[\mathbf{x}_1,\mathbf{x}_2,...,\mathbf{x}_N]\in\mathcal{R}^{D\times N}$ be represented by *d*-dimensional data $\mathbf{Y}=[\mathbf{y}_1,\mathbf{y}_2,...,\mathbf{y}_N]\in\mathcal{R}^{d\times N}$ ($d<<D$), where $\mathbf{y}_i$ denotes a low dimensional representation of $\mathbf{x}_i$. The whole procedure of LLE can be summarized as the following three steps:

- Step 1: Select the k neighbors for each data point $\mathbf{x}_i$, $i = 1,2,...,N$ by pairwise Euclidean distances.
- Step 2: Utilize the neighbors of each data point to compute the linear coefficients (local manifold structure), denoted as $\mathbf{w}_i = [w_i^1,...,w_i^j,...,w_i^k] \in \mathcal{R}^{1\times k}$. The objective function is

$$\mathbf{w}_i = \arg\min_{\mathbf{w}_i} \left\| \mathbf{x}_i - \sum_{j=1}^{k} w_i^j \mathbf{x}_j \right\|^2, \quad \text{s.t.} \sum_{j=1}^{k} w_i^j = 1 \quad (1)$$

- Step 3: Reconstruct the linear coefficients ($\mathbf{w}$) to obtain the embedding in the low-dimensional space.

$$\min \varepsilon(\mathbf{Y}) = \sum_{i=1}^{N} \left\| \mathbf{y}_i - \sum_{j=1}^{k} w_i^j \mathbf{y}_j \right\|^2 \quad \text{s.t.} \sum_{i=1}^{N} \mathbf{y}_i = 0, \ \frac{1}{N} \sum_{i=1}^{N} \mathbf{y}_i \mathbf{y}_i^T = \mathbf{I} \quad (2)$$

From the aforementioned procedures, we can conclude that LLE mainly includes neighbors selection (similarity matrix), local manifold representation and calculation of embedding in the low-dimensional space (spectral decomposition), where construction of a similarity matrix and spectral decomposition are very time-consuming and difficult to be carried out with standard PCs due to its high demand memory and computational cost.

2.2 The proposed K-LLE algorithm

To overcome the aforementioned limits, we propose a new algorithm named K- LLE. K-means clustering is explored to avoid calculating such large similarity matrix and spectral decomposition, making the computation and storage consumption dramatically reduced. The detailed algorithm is described as follows:

- Step 1: Let $\mathbf{C} = [\mathbf{C}_1, \mathbf{C}_2, ..., \mathbf{C}_K] \in \mathcal{R}^{D\times K}$ be K cluster centers with D-dimensional data obtained using K-means clustering. Then, the whole manifold structure of the original data can be approximately represented by K cluster centers. The value of K usually depends on the data size. In this paper, we set the value of K around be 2% of the total number of pixels.
- Step 2: LLE is performed to obtain the d-dimensional representation of K cluster centers by Eq2. (1-2), namely $\mathbf{Y}_C = [\mathbf{y}_{C1}, \mathbf{y}_{C2}, ..., \mathbf{y}_{CK}] \in \mathcal{R}^{d\times K}$. In this way, computational complexity and storage costing is dramatically reduced.
- Step 3: We use K cluster centers to represent local manifold structures instead of all data points and utilize the manifold structure of K cluster centers to represent the whole manifold structure of all data points. As a result, we only need to preserve the local structure between each point and K cluster centers, which is again computationally efficient. Since manifold structure in high-dimensional space is basically consistent with its manifold structure in low-dimensional space [5], we can be assumed that the local manifold structure in the high-dimensional space between a certain data point and K cluster centers is basically consistent with that in the low-dimensional space. Therefore, the low-dimensional representation for each point can be obtained by:

$$\mathbf{y}_c^i = \mathbf{w}_c^i \times \mathbf{Y}_c^i \quad (3)$$

where $\mathbf{y}_c^i$ stands for the low-dimensional representation of i-th data point. $\mathbf{w}_c^i = [w_{c1}, ..., w_{ck}] \in \mathcal{R}^{1\times k}$ is the linear coefficients (local manifold representation) using Eq. (1) between i-th data point and its k nearest cluster centers in the high-dimensional space and $\mathbf{Y}_c^i = [\mathbf{y}_{c1}, ..., \mathbf{y}_{ck}] \in \mathcal{R}^{d\times k}$ is the corresponding k d-dimensional representation selected from $\mathbf{Y}_c = [\mathbf{y}_{c1}, \mathbf{y}_{c2}, ..., \mathbf{y}_{cK}] \in \mathcal{R}^{d\times K}$.

3. EXPERIMENTS

To validate the proposed K-LLE algorithm, we explored classification as a potential application and evaluate the classification accuracy accordingly. One simple but very effective classifier – Nearest Neighbor (NN) based on Euclidean distance, is used for classification. In this paper, we tested the algorithm on two datasets, the AVIRIS Indian Pines dataset with a small size of $145\times145\times220$ and data set from IEEE GRSS Data Fusion Contest 2013 (DFC) with a size of $349\times1905\times144$. In addition, we used the fixed train samples and test samples for two datasets given by [6] and [7], as shown in table 1 and 2, respectively.

Table 1 The number of train samples and test samples for each class in Indian Pine dataset

NO.	Class Name	Total	Training	Testing
1	Corn-Notill	1434	50	1384
2	Corn-Mintill	834	50	784
3	Corn	234	50	184
4	Grass-Pasture	497	50	447
5	Grass-Trees	747	50	697
6	Hay-Windrowed	489	50	439
7	Soybean-Notill	968	50	918
8	Soybean-Mintill	2468	50	2418
9	Soybean-Clean	614	50	564
10	Wheat	212	50	162
11	Woods	1294	50	1244
12	Bldg-Grass-Tree-Drives	380	50	330
13	Stone-Steel-Towers	95	50	45
14	Alfalfa	54	15	39
15	Grass-Pasture-Mowed	26	15	11
16	Oats	20	15	5

Table 2 The number of train samples and test samples for each class in DFC dataset

NO.	Class Name	Total	Training	Testing
1	Healthy Grass	1251	198	1053
2	Stressed Grass	1254	190	1064
3	Synthetic Grass	697	192	505
4	Tree	1244	188	1056
5	Soil	1242	186	1056
6	Water	325	182	143
7	Residential	1268	196	1072
8	Commercial	1244	191	1053
9	Road	1252	193	1059
10	Highway	1227	191	1036
11	Railway	1235	181	1054
12	Parking Lot-1	1233	192	1041
13	Parking Lot-2	469	184	285
14	Tennis Court	428	181	247
15	Running Track	660	187	473

Classification accuracies on the Indian Pine dataset using four methods (original spectral feature (OSF), PCA, LLE, LLE with K-means) under different parameters (the number of neighbors (k), the reduced dimensionality (d)) are shown in Figure 3. Figure 4 shows the classification maps using the optimal parameters. We can see clearly from Figure 3 that the performance of PCA is extremely similar with that of OSF. Significantly, the classification accuracy using K-LLE is superior to that using LLE, which indicates that the proposed method is feasible and effective. In addition, it should be also noticed that LLE and K-LLE outperform OSF and PCA, demonstrating the superiority of manifold learning. Importantly, the classification accuracy obtained by LLE with K-means is relatively stable with the setting of k and d, which implies its robustness. The better performance obtained using K-LLE w.r.t LLE results from that K cluster centers are able to robustly and approximately represent the whole manifold structure of all data points, and hence effectively reduce the influence of data redundancy and a high multicollinearity at each local structure.

Further experiments are performed on the larger-scale hyperspectral dataset. Figure 5 shows the classification accuracies on the IEEE GRSS DFC 2013 dataset and Figure 6 shows the corresponding classification maps under the condition of the best classification accuracies. It can be seen from Figure 5 that K-LLE clearly outperforms the other algorithms. Additionally, as shown in Figure 6, the classification map using K-LLE is relatively more legible than the other classification maps.

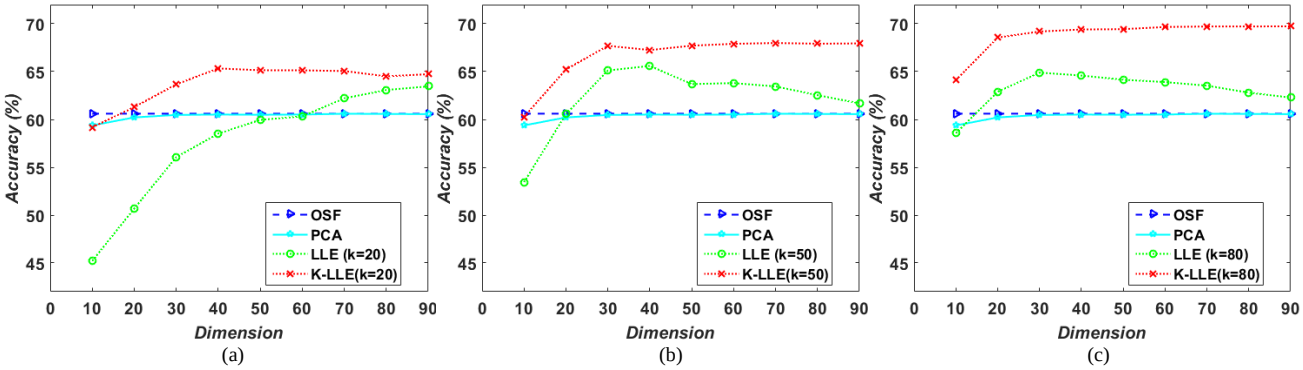


Figure 3 Performance comparison: Classification accuracy as a function of data dimension on the Indian Pine dataset. (a)-(c) are the results using different number of neighbors (k=20, 50, 80) respectively.

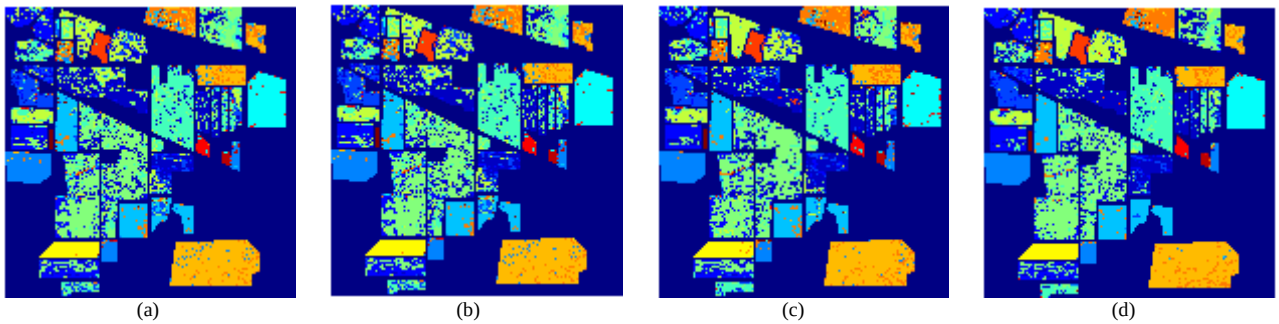
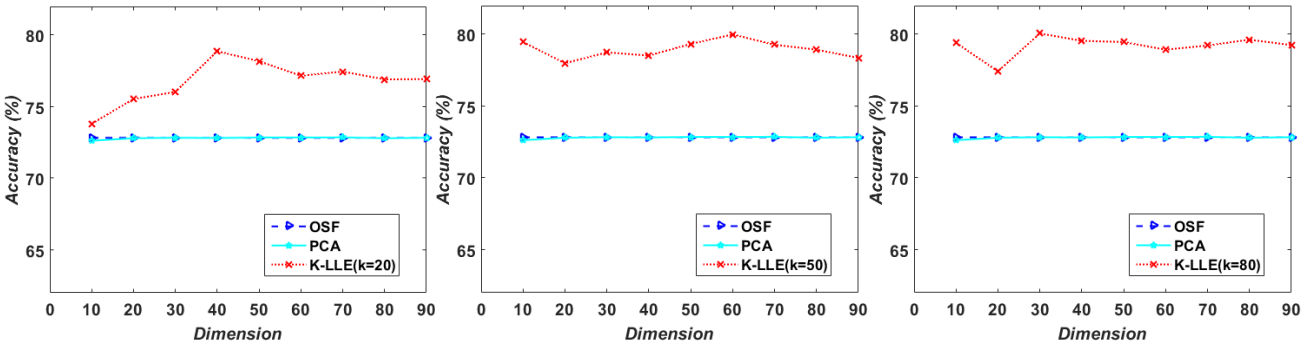
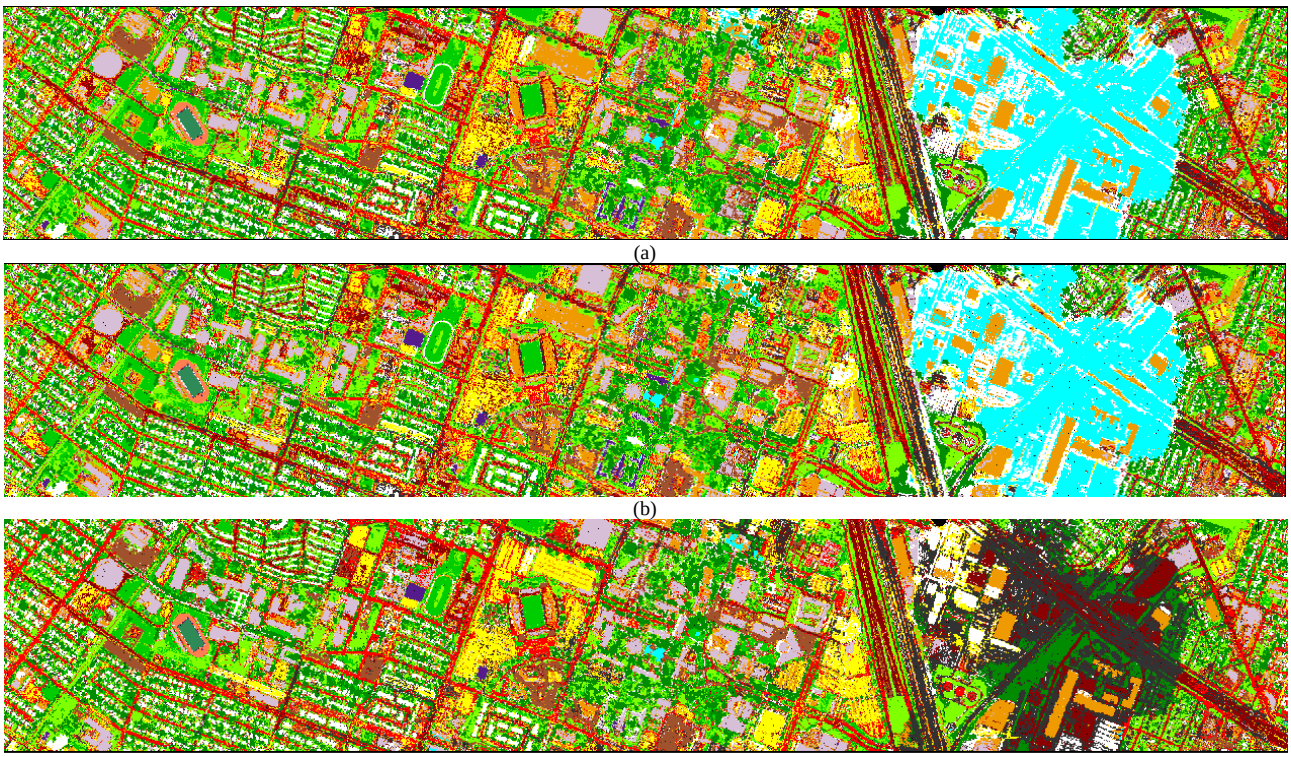


Figure 4 Classification maps using OSF, PCA, LLE, LLE with K-means respectively. (a)-(d) are the results using the optimal parameters for these different methods, respectively (PCA: d=30, LLE: k=50, d=40, K-LLE: k=80, d=90).



(a) (b) (c)

Figure 5 Performance comparison: Classification accuracy as a function of data dimension on the DFC dataset. (a)-(c) are the results using different number of neighbors (k=20, 50, 80) respectively.



(a) (b) (c)

Figure 6 Classification maps using OSF, PCA, and K-LLE respectively. (a)-(c) are the results using the optimal parameters for these different methods, respectively (PCA: d=30, K- LLE: k=50, d=60).

4. CONCLUSION

In this work, we present a method that introduces K-means clustering into LLE to simplify the problem of large-scale spectral decomposition. The proposed K-LLE method enables unsupervised and nonlinear dimensionality reduction in the large-scale hyperspectral data. Experimental results indicate that K-LLE improves classification performance of the original LLE on the different datasets. Also both LLE based methods are superior to PCA and original spectral features. This demonstrates the effectiveness of the proposed strategy. The presented idea can be embedded into any framework that includes spectral decomposition, resulting in the decrease of computational complexity and storage costing. In our future research, we will develop a general framework based on the proposed idea to address the issue of large-scale data in dimensionality reduction, such as manifold learning or kernel learning.

5. REFERENCES

[1] J. B. Tenenbaum, V. de Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," Science, vol. 290, pp. 2319–2323, Dec. 2000.

[2] L. Ma, M. M. Crawford, J. W. Tian, "Local manifold learning-based-k-nearest-neighbor for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sensing*, vol. 48, no. 11, pp.4099-4109, Nov. 2010.

[3] L. Ma, M. M. Crawford, X. Yang, Y. Guo, "Local-manifold-learning-based graph construction for semisupervised hyperspectral image classification," *IEEE Trans. Geosci. Remote Sensing*, vol. 53, no. 5, pp.2832-2844, May. 2015.

[4] A. Talwalkar, S. Kumar, H. Rowley, "Large-scale manifold learning," in *IEEE Computer Society Conference on Computer Vison and Pattern Recognition* (*CVPR*) , pp. 1-8, Jun. 2008.

[5] S. Roweis, L. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol.290, no. 5550, pp.2323-2326, Dec. 2000.

[6] P. Ghamisi, J.A. Benediktsson, M. O. Ulfarsson, "Spectral-spatial classification of hyperspectral images based on hidden markov random fields," *IEEE Trans. Geosci. Remote Sensing*, vol. 52, no. 5, pp.2565-2574, May. 2014.

[7] Y. Zhang, H. L. Yang, S. Prasad, E. Pasolli, J. Jung, M. Crawford, "Ensemble multiple kernel active learning for classification of multisource remote sensing data," *IEEE J. Sel. Topics Appl. Earth Oberve. Romote Sens.*, vol. 8, no.2, pp.845-858, Feb. 2015.