

# LOCAL MANIFOLD LEARNING WITH ROBUST NEIGHBORS SELECTION FOR HYPERSPECTRAL DIMENSIONALITY REDUCTION

Dan Feng Hong <sup>a</sup>, Naoto Yokoya <sup>a,b,c</sup>, Xiao Xiang Zhu <sup>a,b</sup>

<sup>a</sup> Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), Oberpfaffenhofen, 82234 Weßling, Germany,

<sup>b</sup> Signal Processing in Earth Observation (SiPEO), Technical University of Munich, Germany

<sup>c</sup> Research Center for Advanced Science and Technology, The University of Tokyo, Japan

## ABSTRACT

Manifold learning has been successfully applied to hyperspectral dimensionality reduction to embed nonlinear and nonconvex manifolds in the data. However, dimensionality reduction by manifold learning is sensitive to non-uniform data distribution and the selection of neighbors. To address the two issues to some extents, in this work a new manifold framework based on locality linear embedding (LLE), namely local normalization and local feature selection (LNLFS), is proposed. Classification is explored as a potential application to validate the proposed algorithm. Classification accuracy using data obtained using different dimensionality reduction methods is evaluated and compared, while applying two kinds of strategies for selecting the training and test samples: random sampling and region-based sampling. Experimental results show the classification accuracy obtained with LNLFS is superior to state-of-the-art dimensionality reduction methods.

**Index Terms**—hyperspectral image, dimensionality reduction, manifold learning, local normalization, local feature selection, non-uniform data distribution

## 1. INTRODUCTION

Hyperspectral data is characterized by very rich spectral information, which enables us to detect targets of interest, but also introduces drawbacks caused by its high dimensionality. As a result, dimensionality reduction is an important task in hyperspectral data processing. For this purpose, unsupervised manifold learning methods based on a graph embedding framework, such as Isomap [1] and LLE [2], are widely explored [3-5] to embed nonlinear and nonconvex manifolds in hyperspectral data. Neighborhood selection plays an important role in local-graph-based manifold learning. The  $k$ -nearest neighbor algorithm is commonly used for selecting neighbors using distance metrics such as the Euclidean distance, the heat kernel, and the spectral angle distance (SAD). A key limitation of this approach is that it ignores a non-uniform data distribution,

as is always the case in real hyperspectral data. This limitation results in inaccurate calculation of local properties and thus degrades the performance of manifold learning.

This work aims at a robust neighborhood selection method to obtain accurate and robust feature representation via manifold learning. Two main technical contributions proposed in this paper are: 1) an idea of local data normalization has been embedded into the original LLE framework to transform local data points to a more isotropic distribution (i.e. to emphasize spectral features); 2) a novel robust neighbors selection based on local feature matching has been developed to take non-uniform data distributions into consideration.

## 2. METHODOLOGY

### 2.1 Traditional LLE

Figure 1 shows the main procedures of the traditional LLE algorithm, more specific speaking in the following:

Step 1. Select the neighbors for each point in the spectral domain.

Step 2. Utilize the neighbors of each point to describe the local geometric structure, denoted as  $W$ .

Step 3. Preserve the local geometric structure ( $W$ ) in the low dimensional space.

Step 1 plays a very important role in the LLE framework. That is to say, the performance of dimensionality reduction mainly depends on the neighborhood selection. The  $k$ -means algorithm with a distance metric is widely used for the neighborhood selection.

Although the Euclidean distance or the heat kernel is common for calculating the distance between samples [2,4], such distance metrics in the reflectance or radiance domain are highly influenced by spectral variability due to environmental conditions (e.g. illumination and atmospheric conditions) and instrumental configurations (e.g. sensor noise) rather than spectral features. SAD is more suitable for selecting spectrally meaningful neighbors in hyperspectral data while mitigating the effects of variable illumination [3].

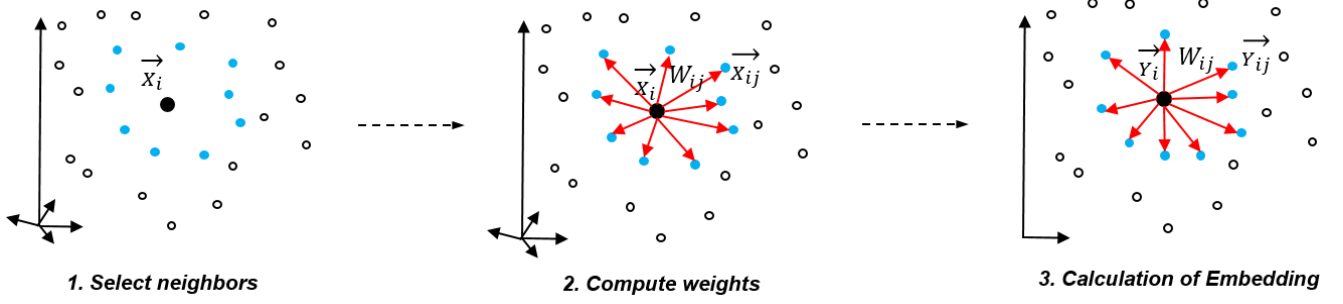


Figure 1 The flowchart of the LLE algorithm

## 2.2 Local Normalization and Local Feature Selection (LNLFS)

We mainly focus on how to select neighbors for each point taking into account non-uniform local data distributions. To this end, we propose a local normalization and local feature selection (LNLFS) method, which fully takes into consideration the non-uniformity of local data distribution and the robustness of local neighbors selection in order to obtain a good representation of local geometric structure.

The idea of data normalization is embedded into the original LLE framework, which can effectively make data distribution more isotropic in the local region, thereby obtain a more accurate representation of local geometric structure.

To handle the locally non-uniformly distributed data and further obtain robust neighbors, we develop an effective neighbors selection method based on local feature matching. It is composed of four core steps: 1) choose a coarse neighborhood for a certain target point in the normalized spectral domain; 2) calculate local features based on the similarity of local manifold structures for all selected neighbors; 3) utilize the KL divergence to measure the similarity with respect to the calculated local features; 4) select a sub-neighborhood with high similarities as the new neighborhood for that target point. Local feature matching and two-step neighborhood selection are inspired by previous works in manifold alignment [6] and clustering [7], respectively.

Compared to strategies of neighbors selections in the previous manifold methods, there are two edges in the new neighborhood obtained with the proposed method as follows: (1) it is based on matching the local features rather than simply measuring a distance metric, therefore is less sensitive to the distribution of original data; (2) it explores a larger neighborhood, and hence makes the local structure calculation more robust. Figure 2 shows the flowchart of the proposed manifold framework and more details corresponding to the Figure 2 are given below.

*Step 1. Select Neighbors:* before neighbors selection, a normalization for the spectral feature is effective to mitigate the effects of variable illumination, making the spectral

feature measured in the same level or unit by using the formula as follows:

$$X_p^N = \frac{X_p - \frac{1}{D} \sum_{j=1}^D (X_{pj})}{\sqrt{\frac{1}{D} \sum_{i=1}^D \left( X_{pi} - \frac{1}{D} \sum_{j=1}^D (X_{pj}) \right)^2}} \quad (1)$$

where  $X_p$  is the spectral signature,  $D$  stands for the number of the spectral bands,  $X_p^N$  is the own normalized spectral feature. Here we use SAD to measure the distance between spectral features and then select the  $K$  nearest neighbors:

$$d_\theta = \cos^{-1} \left( \frac{X_p^N \bullet X_q^N}{\|X_p^N\| \|X_q^N\|} \right) \quad (2)$$

where  $X_p^N$  and  $X_q^N$  are arbitrarily spectral features in the spectral domain, and  $X_p^N \neq X_q^N$ .

*Step 2. Local Normalization (LN):* the idea of normalization is explored before computing the local geometric structure:

$$X_r^{LN} = \frac{\ln X_r^N - \frac{1}{K+1} \sum_{j=1}^K (\ln X_j^N + \ln X_p^N)}{\sqrt{\frac{1}{K+1} \sum_{r=1}^{K+1} \left( \ln X_r^N - \frac{1}{K+1} \sum_{j=1}^K (\ln X_j^N + \ln X_p^N) \right)^2}}, r=1, 2, \dots, K+1 \quad (3)$$

where  $X_p^N$  is the normalized spectral feature itself,  $X_j^N$  stands for the  $K$  nearest neighbors of  $X_p^N$ .  $X_r^N = [X_p^N, X_j^N]$  and  $X_r^{LN}$  can be also written in the form of  $\{X_p^{LN}, X_j^{LN}\}$ ,  $j=1, 2, \dots, K$ , which stands for the local normalized spectral features.

*Step 3. Local Feature Selection:* this step is divided into two parts to describe:

First part: define the local feature  $F_p^{local}$  for the point  $p$  in the spectral domain shown in the following:

$$F_{pj}^{local} = \sum e^{\frac{-(X_p^{LN} - X_j^{LN})^2}{\beta}}, \quad j=1, \dots, K \quad (4)$$

$$F_p^{local} = \{F_{pj}^{local}\}_{j=1}^K \quad (5)$$

Second part: feature matching using KL divergence (KLD)

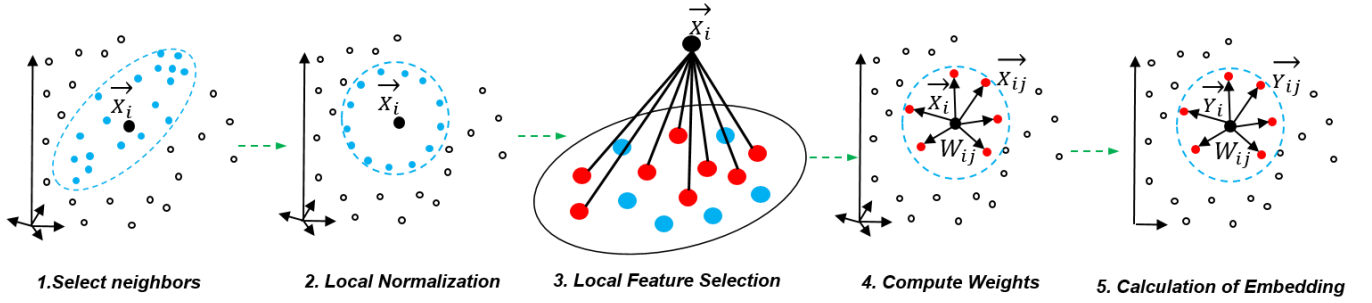


Figure 2 The flowchart of the proposed LNLFS algorithm

$$KLD(\mathbf{F}_p^{local} \parallel \mathbf{F}_j^{local}) = \sum_{i=1}^K \mathbf{F}_p^{local}(i) \times \log_2 \left( \frac{\mathbf{F}_p^{local}(i)}{\mathbf{F}_j^{local}(i)} \right), \quad j=1,2,\dots,K \quad (6)$$

$$KLD(\mathbf{F}_j^{local} \parallel \mathbf{F}_p^{local}) = \sum_{i=1}^K \mathbf{F}_j^{local}(i) \times \log_2 \left( \frac{\mathbf{F}_j^{local}(i)}{\mathbf{F}_p^{local}(i)} \right), \quad j=1,2,\dots,K \quad (7)$$

where  $\mathbf{F}_p^{local}$  and  $\mathbf{F}_j^{local}$  stand for the local features of point  $p$  and the neighbors of point  $p$  in the spectral domain, respectively, and  $\alpha$  is a penalty parameter only to keep a balance between  $KLD(\mathbf{F}_p^{local} \parallel \mathbf{F}_j^{local})$  and  $KLD(\mathbf{F}_j^{local} \parallel \mathbf{F}_p^{local})$ .

So the final matching distance  $\mathbf{d}^f$  between any local features can be expressed as

$$\mathbf{d}^f = \{d_{j=1}^f\}^K = KLD(\mathbf{F}_p^{local} \parallel \mathbf{F}_j^{local}) + \alpha KLD(\mathbf{F}_j^{local} \parallel \mathbf{F}_p^{local}) \quad (8)$$

Then choose  $k$  neighbors with the small  $\mathbf{d}^f$  value from the  $K$  nearest neighbors. Therefore, the new neighbors  $\mathbf{W}_p$  for certain point can be updated using the formula as follows:

$$\mathbf{W}_p = \{\mathbf{W}_{ij}\}_{j=1}^k = \min_{j=1}^k \left( \{d_{j=1}^f\}^K \right) \quad (9)$$

Step 4. Compute Weights:

$$\min \varepsilon(\mathbf{W}_i) = \left| \mathbf{X}_i^{LN} - \sum_{j=1}^k \mathbf{W}_i^j \mathbf{X}_{ij}^{LN} \right|^2, \quad s.t. \sum_{j=1}^k \mathbf{W}_i^j = 1 \quad (10)$$

Step 5. Calculation of Embedding:

$$\min \varepsilon(\mathbf{Y}) = \sum_{i=1}^N \left| \mathbf{Y}_i - \sum_{j=1}^k \mathbf{W}_i^j \mathbf{Y}_{ij} \right|^2, \quad s.t. \sum_{i=1}^N \mathbf{Y}_i = 0, \quad \frac{1}{N} \sum_{i=1}^N \mathbf{Y}_i \mathbf{Y}_i^T = \mathbf{I} \quad (11)$$

where the solution  $\mathbf{Y}$  is the representation of dimensionality reduction for the original hyperspectral data.

### 3. EXPERIMENT

In order to validate the performance of LNLFS, we explored classification as a potential application and evaluate the classification accuracy accordingly. Here, we adopted two sampling strategies to select training samples and test samples: one is random sampling which is a common way for the validation of hyperspectral classification and another is a region-based sampling which is more challenging as well as an effective way to check the robustness of the proposed method. In addition, we used the AVIRIS Indian Pines dataset (200 bands without water absorption bands) for the experiments mentioned above. The Indian Pines dataset has 16 classes with greatly different numbers of total

samples as shown in the Table 1. The data include nonlinear and nonconvex manifolds with non-uniform data distribution due to multiple scattering, bidirectional reflectance distribution function effects, variable illumination conditions, and sensor noise. Therefore the Indian Pines is a suitable dataset to evaluate the performance of our method.

Table 1 The number of train samples and test samples for each class

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

Figure 3 and Figure 4 shows the classification accuracy using the different methods including original spectral feature, principle component analysis (PCA), LLE with the Euclidean distance [4], Laplacian Eigenmaps (LE), local normalization (LN), and LNLFS, under the different sampling strategies. In addition, we only use the 1 nearest neighbor based on the Euclidean distance as a classifier. We can see clearly from the Figure 3 that in the dataset with non-uniform data distribution and variability, LN and LNLFS methods obtain the better classification results compared to the original spectral feature and LLE and LE, and LNLFS is obviously superior to the LN as well. However, LLE and LE are sensitive to the variability of data. Due to the spectral variability in the data, such as noise and shadow, etc., their performances are even inferior to those using directly original spectral feature and PCA. Furthermore, in order to validate the effectiveness and robustness, we used another strategy based on the region to

obtain training samples for classification. As shown in the Figure 4, our method LN and LNLFS are still superior to other methods, though their classification accuracy degrades as expected. In addition, we can also notice from the figures that the classification accuracy using PCA based on linear dimensionality reduction is extremely similar with that directly using the original spectral feature under the different

sampling strategies, which means that there is not an effective improvement in performance after using linear dimensionality reduction method, while the proposed method based nonlinear has better performance, which is more suitable for the hyperspectral data dimensionality reduction due to the special nonlinear property of the hyperspectral data.

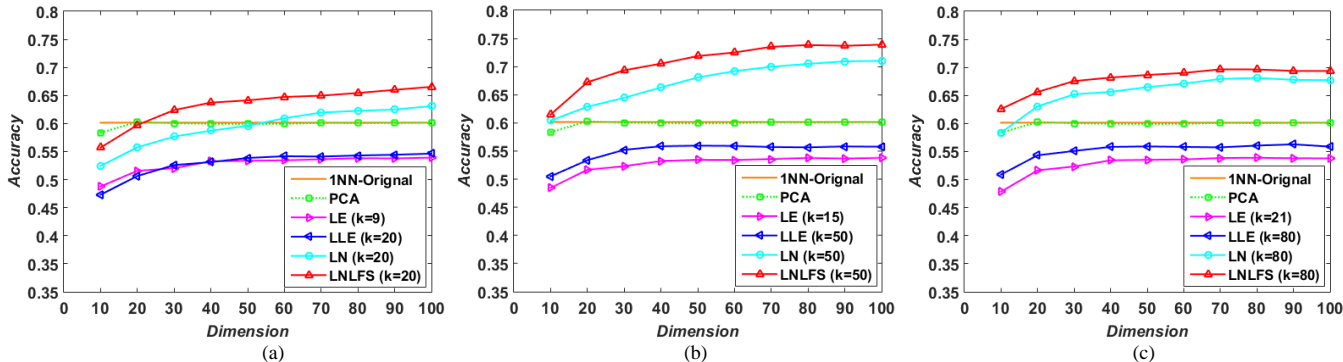


Figure 3 Performance comparison: Classification accuracy as a function of data dimension using randomly sampling (our proposed method: LNLFS).

(a)-(c) are the results using different number of neighbors (k=20, 50, 80) respectively.

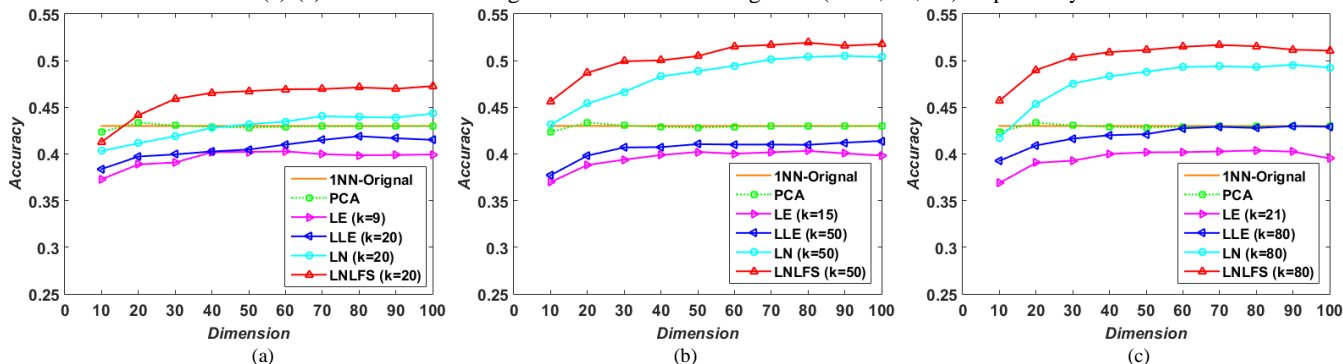


Figure 4 Performance comparison: Classification accuracy as a function of data dimension using region-based sampling (our proposed method: LNLFS).

(a)-(c) are the results using different number of neighbors (k=20, 50, 80) respectively.

#### 4. CONCLUSION

In this work, a novel neighborhood selection method for LLE – the LNLFS method – is developed to handle non-uniform data distribution in nonlinear and nonconvex manifolds of hyperspectral data for dimensionality reduction. Compared to other state-of-the-art methods, the proposed method achieves better performance in terms of classification accuracy. In the future, we will further focus on how to better describe the local geometric structure and how to automatically obtain the number of neighbors, thereby try to develop a more automated framework.

#### 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] S. Roweis, L. Saul, “Nonlinear dimensionality reduction by locally linear embedding,” *Science*, 290 (5550), pp.2323-2326, 2000.

[3] C. M. Bachmann, T. L. Ainsworth, and R. A. Fusina, “Exploiting manifold geometry in hyperspectral imagery,” *IEEE Trans. Geosci. Remote Sensing*, vol. 43, no. 3, pp. 441–454, 2005.

[4] 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, 2010.

[5] D. Lungu, S. Prasad, M. M. Crawford, and O. Ersoy, “Manifold-learning-based feature extraction for classification of hyperspectral data,” *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 55-66.

[6] Y. Pei, F. Huang, F. Shi, H. Zhi, “Unsupervised image matching based on manifold alignment,” *IEEE Trans. Pattern Anal. Mach. Intell.*, 34 (8), pp.1658-1664, 2012.

[7] C. Wang, J. Lai, J. Zhu, “Graph-based Multiprototype Competitive Learning and its applications,” *IEEE transactions on Systems, Man, and Cybernetics-Part C*, 42 (6), pp.934-946, 2012.