Nonlinear Feature Extraction for Hyperspectral Images

Cigdem BAKIR*

Accepted 15th August 2014 DOI: 10.18100/ijamec.74610

Abstract: In this study non-linear dimension reduction methods have been applied to a hyperspectral image in order to increase the classification accuracy in feature extraction step. Furthermore, image segmentation has been ensured by taking into consideration the spatial synthesis of hyperspectral images and passing from high-dimensional space to low dimensional space. It has been compared the results obtained from the image segmentation made by taking one pixel from this spatial synthesis. The advantages of the effects of the results of the dimension reduction techniques made by facing neighbor pixels on the segmentation of super-spectral image have been displayed in the experimental results part.

Keywords: Feature extraction; Spatial dependency; Dimension reduction; Manifold learning.

1. Introduction

With the advance of remote sensing technology in the recent years hyperspectral image scanners have become very popular in many scientific area such as geosciences and medicine. If hyperspectral images are compared with multispectral images, hyperspectral images contain much richer data. However, the presence of hundreds of adjacent spectral bands in the hyperspectral images negatively affects the pattern recognition algorithm. That is why the most considerable problem of the hyperspectral image processing is the dimension reduction and feature extraction. Feature extraction and dimension reduction techniques are very relevant to each other in their nature. By extraction of new features from original spectral bands as a parallel process of dimension reduction, classification and segmentation algorithms can be used more effectively from machine learning and pattern recognition area in a lower dimensional space.

Dimension reduction methods can be realized in two ways: feature selection or feature extraction. In feature selection methods only a small subset of highly informative bands are chosen. Feature extraction refers to the mapping of high dimensional data onto a lower dimensional space. In feature extraction methods features are extracted using linear or nonlinear functions of the original bands.

In this study dimension reduction, band selection and especially characteristic inference is done. Furthermore, by benefiting from the characteristic inference from hyperspectral images and using class labels with informative methods, it is tried to perform classification analysis. There are many methods to execute the characteristic inference. The oldest and hence the most known of these techniques is the linear technique named PCA. However, linear techniques are not sufficient in data inference. It is needed to have non linear techniques to classify complex data. In the recent years, many non linear techniques have been used for this purpose. In addition to the non linear techniques protecting global characteristics such as Kernel PCA, Isomap, Diffusion Maps we have also used non-linear techniques protecting local characteristics such as LLE, Hessian LLE, Laplacian Eigenmaps, LLC. [9] Furthermore, it has used PCA to compare non-linear techniques.

As second step, it have been obtained the 5, 10, 15, 20 and 50 spatial neighborhoods for the available data set created with this characteristic inference methods and tried to classify the images. Each data is expressed as a sample point. This sample point is named as pixel. The vectors created for these pixels are valid for all bands and have a specified value. Is the class of a pixel is determined; the class of its neighborhood next to it is the same and represents its spatial neighborhood [1]. This spatial neighborhood for a pixel of an image is to take n×n neighborhood around it and to discipline to image. The proximity of spatial neighborhood between two points is measured with Euclidean distance. In spatial neighborhoods, two close data points are similar and their probability to be in the same class is very high. As distance increases, similarity decreases for n×n neighbor.

In the second part, after displaying the linear and nonlinear techniques, the data set has been presented as well the study in the third part. In the last part, the experimental results have been showed as well as advantages of this study. Classification performance can be kept same or can be increased.

2. Methodology

In the real world, it is been struggled too much data. It is needed too much memory, calculation to correctly classify the data. With the reduction of the data it is possible to use it efficiently and to reduce cost. And to perform this operation without any data loss, you need dimension reduction methods ensuring dimension reduction and decreasing band number. This is specially an important step for the processing of high inference dimensional data. The characteristic inference is done on bands with linear or nonlinear methods. In this study it has been done characteristic inference by using class data[14]. In the following method of dimensionality reduction used in this study are presented.

2.1. Diffusion Maps

Diffusion Maps technique is based on Markov chain base. The diffusion distance between two points x and y is known as casual
movement probability. This range possibility is identified as follows:

\[
D^{ij}(x_i, x_j) = \sqrt{\sum_k (p_{ik}(t) - p_{jk}(t))^2 / \sqrt{x_{ik}(0)}^2}\]  \hfill (1)

In Equation (1) \( \phi(x_i) \) shows the over-weighted characteristics. Diffusion Maps and end weight in data graphic are calculated with Gauss kernel function[2]. This method combines all along the graph all ways and uses diffusion distance to reduce dimension. It is continued to process all along the graph by throwing minimum eigenvalue and eigenvector.

2.2. Kernal PCA

Kernel PCA-KPCA is the expanded feature of PCA method. However, while PCA is a linear method KPCA is a non-linear technique that improves linear techniques. Data dimension is reduced by using Kernel matrix, so as K kernel matrix of the data points \( x_i \) in the form of \( K(x_i, x_j) \). By changing K kernel matrix inputs, it is possible to find the centers and d eigenvectors. The essential in KPCA is to select the kernel function. Kernel function can be linear kernel, Gauss kernel and polynomial kernel. KPCA ensures quite successful results in face detection, speech recognition.

2.3. Principal Component Analysis (PCA)

PCA is the most popular orthogonal linear transformation. PCA is display of the data having the biggest variance in low dimensional space. High variance characteristics are preferred to low variance. By calculating to covariance matrix of X data matrix samples, it is possible to find the M linear mapping increasing the cost function. It removes the eigenvectors having the biggest eigenvalue. PCA used the euclidean distance between \( x_i \) and \( x_j \) data points. PCA transformation is as \( \mu^T=W \) orthogonal matrix, \( \mu^T \) linear transformation; whereas \( W \) displays the eigenvector corresponding to covariance matrix.

2.4. Isomap

Instead of Euclidean distance used in multidimensional scaling algorithm on Isomap weighted graphic, it is a low dimensioned embedding method using geodesic distance. Geodesic distance is the shortest one and it is needed to find neighborhoods between all data points in order to find this distance. Geodesic distance is calculated with neighborhood graph between data points \( x_i (i=1,2,...,n) \). Furthermore, in Isomap method it is also important to select the neighborhood parameter. The linkage of each data point is known as the closest Euclidean distance in high dimension space. The shortest distance between two points is found with Dijixtra algorithm.

2.5. Local Linear Embedding (LLE)

It is a method similar to Isomap. However as distinct from Isomap, it only protects the local characteristics on data point graphic image. In LLE \( x_i \) data point converging to local neighbors and \( k \) \( w_i \) weights are matching a hyper smoothing being a linear union of their closest neighbors by preserving the neighbor relations of each data point. \( w_i \) weights are stable against reconfiguration, transformation, slewing, scaling. Various coefficients are created. \( W=(W_{ij})_{i,j} \) minimizes the cost function. After calculating the weights, it is passed to low dimensioned space by preserving local neighbors.

2.6. Hessian LLE

As LLE discrete matrix techniques are used. It calculates the data graphic by using k the closest neighborhood. It performs preliminary analysis by measuring manifold curves of H matrix and displays in low dimensioned space. It is calculated as Tangent Hessians:

\[
H_{lm} = \sum_i ((H_i)_l)^*((H_i)_m)
\]  \hfill (2)

For each data point, approximate local tangent coordinates are calculated for manifold Hessians by calculating eigenvectors of covariance matrix. The H matrix is performed minimizes curviness of the Afterwards, the matrix created in order to determined eigenvectors of covariance matrix is orthonormalized. Compared to LLE and Laplacian it is slowest and gives worst performance in discrete ones. However it is successful in convergence problems.

2.7. Laplacian Eigenmaps

As similar to LLE, it passes to low dimensioned space by preserving manifold local characteristics. Laplacian method creates first G graphic in association with k the closest neighborhood and selects the weight. The end weights for \( x_i \) and \( x_j \) data points are calculated with gauss. \( y_i \) is subjected to \( x_i \) weights in cost function of low dimensioned space image and little distances between \( x_i \) data points. Hence, cost function is minimized with \( y_i \) and \( y_j \) spectral graphic theory. M is stated as grade matrix, \( W \) diagonal matrix and Laplacian graphic L=M-W. Eigen decomposition of Laplacian graphic is done and low dimensioned embedding is created.

2.8. Locally Linear Coordination (LLC)

LLC performs global editing of linear models by calculating linear models. A mixture of factor analysis is calculated with LLC EM(Expectation Maximization) algorithm. The mixture of factor m gives local data images and \( r_{ij} \), \( y_{i,j} \) which are the related responsibilities. \( W \) weight matrix shows the weights created with LLE by using \( r_{ij} \). The weights are stated as:

\[
y_{i,j} = \sum_{i=1}^{k} r_{ij} (L_{ij} + o_i)
\]  \hfill (3)

2.9. Multidimensional Scaling (MDS)

MDS is a technique based on similarity matrix. It matches as much as possible the points having closest original distance for similarity between N couples of data. This indicates how much the objects resemble to each other. Furthermore, MDS finds also matrix distances that are not like. There are many MDS algorithms and classifies these MDS algorithms depending on entry matrix.

3. Experimental Study and Results

In this study Salinas data has been used set obtained from 224 bands captured from AVIRIS (Airborne Visible/Infrared Imaging Spectrometer) sensor. Salinas data set is a data set offering high extensional resolution found in the valley of Salinas in California. It is composed of 217 samples, 512 lines and in whole 16 classes. In this study, in order to ensure at first the processing of all data, it has been obtained the signatures ensuring some characteristics. It is thought that these spectral signatures represent as far as possible the spectral range. The non-linear techniques allowing characteristic inference have been applied not to all data but to these spectral signatures. The data set was obtained by removing 10 and 15 featured vectors thanks to dimension reduction techniques used in this study. Furthermore, RBF and kNN interpolation was used in order to discipline all data set. The data set reduced with RBF was learnt with artificial neural networks and...
each spectral signature passed to low dimensioned space by using a network structure. kNN interpolation method is found according to k closest samples. In kNN interpolation, weights are calculated by finding the closest neighborhood to normalized image. In this method, spectral signatures and distances of each pixel are found. There after by selecting k=9 value feature vectors are obtained by getting nonlinear dimension reduction method of k spectral signature which is closest to pixel to be reduced. The purpose here is to find the closest mid evaluation to be obtained with non-linear projection methods. This method can be applied on all spectral signatures and ensures dimension reduction.

The experimental results for Salinas data set are for 10, 15 dimensions. Salinas data set was reduced to 10, 15 bands by using 224 bands extensional union. Segmentation and classification success were measures by using 5, 10, 15, 20 and 50 neighborhoods. Segmentation results obtained by using one pixel as well as segmentation results obtained by taking neighborhoods for 16 different classes were evaluated. Have also displayed the results obtained from a linear method PCA for comparison purpose. The results of the study are shown in the following table. In the Table I and Table II have been displayed RBF results for 10 and 15 dimension respectively. In the Table III have been displayed RBF results obtained by ensuring spatial union for 10 dimension. With modified non-linear embedding, it is obtained a classification the highest accuracy of 99.51% by 10 bands. According to classical RBF have been increased by about 10%. In the Table IV have been displayed RBF results obtained by ensuring spatial union for 15 dimension. The highest accuracy is obtained with modified and with recommendation non-linear embedding. Recommended technique in all non-linear classification techniques is increased success.

Table 1. RBF results for 10,15 dimensions

| Dimension Reduction Methods | RBF | 10 dimension | 15 dimension |
|----------------------------|-----|--------------|--------------|
| None reduction             |     | 68.88        | 68.89        |
| Diffusion Maps             |     | 84.39        | 88.69        |
| KPCA                       |     | 69.25        | 69.56        |
| LLC                        |     | 50.2         | 51.28        |
| PCA                        |     | 69.12        | 68.59        |
| LLE                        |     | 51.2         | 51.28        |
| Isomap                     |     | 68.09        | 66.51        |
| MDS                        |     | 69.12        | 68.59        |
| Laplacian E.               |     | 51.23        | 51.28        |
| Hessian LLE                |     | 51.2         | 51.3         |

Table 2. kNN results for 10,15 dimensions

| Dimension Reduction Methods | kNN (k=9) | 10 dimension | 15 dimension |
|-----------------------------|-----------|--------------|--------------|
| None reduction              |           | 65.73        | 65.73        |
| Diffusion Maps              |           | 76.88        | 78.58        |
| KPCA                        |           | 68.22        | 68.27        |

Table 3. RBF Spatial Neighborhood Results for 10 Dimension

| Dimension Reduction Methods | RBF(10 dimension) |
|-----------------------------|-------------------|
|                            | 5     | 10     | 15     | 20     | 50     |
| Diffusion M.               | 84.18 | 96.13 | 99.51  | 99.38  | 66.33  |
| KPCA                       | 70.23 | 93.34 | 99.03  | 95.77  | 50.67  |
| LLC                        | 62.85 | 78.86 | 79.25  | 74.25  | 45.65  |
| PCA                        | 77.14 | 88.87 | 90.63  | 90.15  | 65.36  |
| LLE                        | 70.2 | 83.52 | 88.65  | 84.12  | 50.32  |
| Isomap                     | 65.78 | 89.53 | 89.63  | 84.36  | 50.78  |
| MDS                        | 79.58 | 98.63 | 99.15  | 97.56  | 69.63  |
| Laplacian E.               | 65.17 | 80.58 | 82.18  | 83.63  | 45.67  |
| Hessian LLE                | 70.25 | 82.29 | 87.52  | 83.29  | 45.28  |

Table 4. RBF Spatial Neighborhood Results for 15 Dimension

| Dimension Reduction Methods | RBF(15 dimension) |
|-----------------------------|-------------------|
|                            | 5     | 10     | 15     | 20     | 50     |
| Diffusion M.               | 79.81 | 97.77 | 99.13  | 98.83  | 65.1   |
| KPCA                       | 65.38 | 92.78 | 98.72  | 96.52  | 42.31  |
| LLC                        | 60.33 | 71.23 | 70.22  | 71.36  | 44.36  |
| PCA                        | 76.85 | 85.14 | 89.97  | 88.69  | 62.81  |
| LLE                        | 68.87 | 82.41 | 83.2   | 80.1   | 45.87  |
| Isomap                     | 66.72 | 85.63 | 88.11  | 86.93  | 48.36  |
| MDS                        | 75.69 | 96.78 | 96.36  | 95.36  | 65.66  |
| Laplacian E.               | 68.65 | 71.54 | 75.64  | 73.53  | 42.14  |
| Hessian LLE                | 70.25 | 82.29 | 87.52  | 83.29  | 45.28  |

Table 5. kNN Spatial Neighborhood Results for 10 Dimension

| Dimension Reduction Methods | kNN(10 dimension) |
|-----------------------------|-------------------|
|                            | 5     | 10     | 15     | 20     | 50     |
| Diffusion M.               | 77.75 | 86.81 | 99.39  | 97.14  | 49.73  |
| KPCA                       | 68.51 | 96.58 | 98.12  | 95.32  | 47.63  |
| LLC                        | 58.78 | 71.25 | 74.1   | 72.25  | 40.22  |
In the Table V have been displayed kNN results obtained by ensuring spatial union for 10 dimension. The highest accuracy is obtained with 15 and 20 spatial neighborhoods of KPCA, Diffusion Maps techniques. The success is by about 99%. In the Table VI have been displayed kNN results obtained by ensuring spatial union for 15 dimension. The highest accuracy is obtained with 15 and 20 spatial neighborhoods of Diffusion Maps, KPCA and MDS techniques. PCA which linear technique has given very good results in the classification.

### Table 6. kNN Spatial Neighborhood Results for 15 Dimension

| Dimension Reduction Methods | 5     | 10    | 15    | 20    | 50   |
|-----------------------------|-------|-------|-------|-------|------|
| Diffusion M.                | 79.88 | 82.66 | 98.34 | 95.67 | 53.14|
| KPCA                        | 77.56 | 94.52 | 97.93 | 96.22 | 40.58|
| LLC                         | 52.14 | 70.32 | 73.72 | 71.23 | 34.23|
| PCA                         | 74.22 | 87.77 | 88.37 | 84.37 | 64.77|
| LLE                         | 60.36 | 82.21 | 83.52 | 84.58 | 40.26|
| Isomap                      | 60.97 | 85.14 | 87.34 | 86.15 | 65.17|
| MDS                         | 74.33 | 89.65 | 90.53 | 92.36 | 67.31|
| Laplacian E.                | 65.96 | 78.36 | 75.55 | 76.36 | 45.33|
| Hessian LLE                 | 64.25 | 84.98 | 86.22 | 84.24 | 47.77|

### 4. Conclusion

In this study, it has introduced spatial coherence into dimensionality reduction and showed its contribution for the image segmentation. The spatial coherence is introduced by comparing individual pixels based on neighborhood structure of their non-linear dimension reduction techniques. Furthermore, using spatial coherence have been tried to increase the classification success of dimension reduction techniques. Reduction to 10 bands and 15 bands considering 10-20 neighbors gave the best classification results. Compared to results the classification accuracy decreased when the reduced too few of band or too many of bands. It is important to express the best by taking the appropriate number spatial neighborhood.

### References

[1] A. Mohan, G. Sapiro and E. Bosch, “Spatially coherent Nonlinear Dimensionality Reduction and Segmentation of Hyperspectral Images,” IEEE Geosci.and Remote Sens., vol. 4, no.2, pp.206-210, April 2007.

[2] L. Macter, E. Postma, and J. Heik, “Dimensionality reduction:A comparative review”, Ticc Tilburg University, October 26, 2009.

[3] E. Namey, G. Guest, L. Thairy and L. Johnson, “Data reduction techniques for large qualitative data sets”, Handbook for team-based qualitative research, pp.137-162, March 29, 2007.

[4] Y. Hou, P. Zhang, X. Xu, X. Zhang and W. Li, “Nonlinear dimensionality reduction by Locally Linear Inlaying,” IEEE Transactions of Neural Networks, vol.20, no.2, pp.300-315, February 2009.

[5] W. Kim, M. Crawford and S. Lee, “Integrating spatial proximity with manifold learning for hyperspectral data,” Korean Journal of Remoting Sensing, vol.26, no.6, pp.693-703, 2010.

[6] Melgani F, M. Bruzzone L., “Classification of hyperspectral remote sensing images with support vector machines,” IEEE Transactions Geoscience and Remote Sensing on volume 42, Issue 8, pp 1778-1790, Agust 2004.

[7] X. Huang, L. Zhang, “A comparative study of spatial approaches for urban mapping using hyperspectral Rosis images over Pavia City,” volume 30, Issue 12, 2009.

[8] M. Fong, “Dimension Reduction on Hyperspectral Images”, Project Report, REU projects, UCLA Dept of ath. Agu 2007.

[9] Melba M. Crawford, Li Ma and W. Kim, “Exploring nonlinear manifold learning for classification of hyperspectral data,” Augmented vision and reality volum 3, pp 207-234, 2011.

[10] Kim W, “Multiresolution manifold learning for classification of hyperspectral data,” IEEE International Geoscience and Remote Sensing Symposium, pp.3785-3788, July 2007.

[11] M. Fauvel, J. Benediksson, J. Chanussot and Johannes R. Sveinsson, “Spectral and spatial classification of hyperspectral data using SVMs and morphological profiles,” IEEE Transactions on Geoscience and Remote Sensing Society,vol.46, no.11, pp.30804-3814, Nov 2008.

[12] Y Chen, M.Crawford and J. Ghost, “Applying nonlinear manifold learning to hyperspectral data,” Augmented vision and reality vol.6, pp.4311-4314, 2005.

[13] S.T.Roweis and L.K.Saul, “Nonlinear dimensionality reduction by Local Linear Embedding,” Science, vol.290, pp.2323- 2326, 2000.

[14] Shen-En Qian, Chen G, “A new nonlinear dimensionality reduction method with application to hyperspectral image analysis,” IEEE International Geoscience and Remote Sensing Symposium,pp.270-273, July 2007.

[15] Charles M.Banchmann, Thomas L.Ainsworth, Robert A.Fusina, “Exploiting manifold geometry in hyperspectral images,” IEEE Transactions on volume 43, No 3, pp.441-454, March 2005.

[16] Bo Du, L. Zhang, L. Zhang, T. Chen and Ke Wu, “A discriminative manifold learning based dimension reduction method for hyperspectral classification,” International Journal of Fuzzy Systems, vol.14, no.2, June 2012.

[17] Costa J.A, Hero A.O,“Classification constrained dimensionality reduction,” IEEE International Conference on Speech and Signal Processing, vol 5, pp. 1077-1080, March 2005.

[18] Wang X.R, Ramos F, Kaupp T,“Probabilistic classification of hyperspectral images by learning nonlinear
dimensionality reduction mapping,” International Conference on Information Fusion, July 2006.

[19] L. O. Jimenez, and D. A. Landgrebe, “Supervised Classification in High Dimensional Space: Geometrical, Statistical, and Asymptotical Properties of Multivariable Data,” IEEE Transactions on Systems, Man and Cybernetics, vol 28, no. 1, pp. 39-54, Feb. 1998.

[20] 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.