Application of Curve Fitting in Hyperspectral Data Classification and Compression

Seyed Abolfazl Hosseini
Department of Electrical and Computer Engineering, Tarbiat Modares University, Tehran, Iran
abolfazl.hosseini@modares.ac.ir
Hassan Ghassemian*
Department of Electrical and Computer Engineering, Tarbiat Modares University, Tehran, Iran
ghassemi@modares.ac.ir

Received: 10/May/2015 Revised: 24/Jun/2015 Accepted: 30/Jul/2015

Abstract

Regarding to the high between-band correlation and large volumes of hyperspectral data, feature reduction (either feature selection or extraction) is an important part of classification process for this data type. A variety of feature reduction methods have been developed using spectral and spatial domains. In this paper, a feature extracting technique is proposed based on rational function curve fitting. For each pixel of a hyperspectral image, a specific rational function approximation is developed to fit the spectral response curve of that pixel. Coefficients of the numerator and denominator polynomials of these functions are considered as new extracted features. This new technique is based on the fact that the sequence discipline - ordinance of reflectance coefficients in spectral response curve - contains some information which has not been considered by other statistical analysis based methods, such as Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) and their nonlinear versions. Also, we show that naturally different curves can be approximated by rational functions with equal form, but different amounts of coefficients. Maximum likelihood classification results demonstrate that the Rational Function Curve Fitting Feature Extraction (RFCF-FE) method provides better classification accuracies compared to competing feature extraction algorithms. The method, also, has the ability of lossy data compression. The original data can be reconstructed using the fitted curves. In addition, the proposed algorithm has the possibility to be applied to all pixels of image individually and simultaneously, unlike to PCA and other methods which need to know whole data for computing the transform matrix.

Keywords: Hyperspectral; Feature Extraction; Spectral Response Curve; Curve Fitting; Classification.

1. Introduction

Hyperspectral (HS) data contain hundreds of spatially coregistered images in the form of image cubes; each image corresponds to a specific narrow spectral band usually in the wavelength range of 400-2500 nm. In other words, pixels of an N bands HS data set can be considered as N-dimensional vectors containing sequential intensities that are measured by the HS sensor. For each pixel, the plot of bands intensity values ($y= [y_1, y_2 \dots y_N]^t$) vs. band numbers ($x=[1, 2,\dots, N]^t$) is named as spectral signature or spectral response curve (SRC) of pixel. Some theoretical and practical problems appear in supervised classification of this type of data because of high dimensional spaces specifications. The most important problem in is Hughes phenomenon [1]. There are four strategies to combat Hughes phenomenon in hyperspectral images classification: semi-supervised classification [2,3], combination of spatial and spectral information [4-6], utilizing classifiers like SVM which are less sensitive to the number of training samples [7], and feature reduction (extraction/selection) [1].

Statistical analysis based transformations for performing unsupervised or supervised feature reduction approaches are powerful methods in classification and visualization of remote sensing data [8,9]. In feature reduction we find a transformation that maps data to a lower dimensional space, mainly preserving essential discriminative information. Great efforts have been made to develop advanced feature reduction and classification methods to improve classification accuracy rate. Principal component analysis (PCA) [1], decision boundary feature extraction (DBFE), non-parametric weighted feature extraction [10], wavelet transform [11], maximum margin projection (MMP) [12], linear discriminant analysis (LDA) [13], and independent components analysis (ICA) [14] are examples of feature extraction (F.E.) techniques for data redundancy reduction in remotely sensed data. Nonlinear extensions of PCA (KPCA) have been proposed by using kernel trick [15]. Kernel Fisher discriminant analysis (KFD) [16] and Generalized discriminant analysis (GDA) [17] have been developed independently as kernel-based nonlinear extensions of LDA. Also optimal selection of spectral bands has been extensively discussed in the literature by utilizing approaches like sequential forward/backward methods and information theoretical based methods [1]. After feature reduction step, the features are fed to the classifier. A widely used parametric classifier in hyperspectral data classification is the maximum likelihood classifier [18].

* Corresponding Author

Although methods like PCA, LDA and many other statistical analysis based methods have simple structure and relatively good results, but there are some deficiencies when they are used in hyperspectral Image classification. An important deficiency is that these methods do not consider the geometric aspects of SRCs and the ordinance of original features which is a rich source of information. For each pixel of a hyperspectral image we have a vector of measured quantities corresponding to reflection coefficients of consecutive wavelengths (SRC). Therefore the ordinance of measured data might have some information that could be useful in classification process. In another word, SRC, as a geometrical curve has some useful information. A feature extraction techniques based on the fractal nature of SRC was introduced by [19,20] that depends on the ordinance of samples. Other disadvantage of many F.E. methods is that they cannot be applied to individual pixels independently and first we need to transform whole data to new space and then produce new features. The main contribution of the present letter is introducing another F.E. method which considers the geometrical nature of the SRCs and the ordinance information existing in the SRC that yields improvement of correct classification rate and some other evaluation factors. Indeed we try to fit a rational function with polynomial numerator and denominator to the SRC of each pixel. Then the coefficients of these polynomials are considered as new feature vectors and are fed to an ML classifier. Results are compared to PCA and LDA as two basic and classic F.E. methods with relatively good performance for HS data classification. Unlike PCA and LDA, our proposed method is applied pixel by pixel and does not need to transform whole data to a new space. Therefore, a parallel implementation of the algorithm is possible. Moreover, since the proposed transform is invertible, it could be used as a lossy compression algorithm for HS data, too.

2. Curve Fitting and its Discriminating Ability

In many measurement problems, curve fitting is a traditional approach to find the mathematical relationship between observed values and independent variables. It tries to fit an appropriate curve to the observed values. In addition, curve fitting can be used for noise reduction and data smoothing purposes, and data interpolation / extrapolation [21,22]. The aim of curve fitting is to find a function $f(\lambda)$ in a pre-specified class of functions for the data $\{(\lambda, I_\lambda)\}$ where λ= 1,2,…,N, which minimizes the residual (the distance between the data samples and $f(\lambda)$) under the weight W= (w_λ) [23].

Least squares (LS), least absolute residuals (LAR), and bisquare fitting method are some of fitting criteria used to perform linear or nonlinear fittings to find the function $f(\lambda)$ [24]. In the LS method $f(\lambda)$ is found by minimization of the following weighted mean squared error:

$$\frac{1}{N}\sum_{\lambda=1}^{N} w_\lambda (f(\lambda) - I_\lambda)^2 \qquad (1)$$

There are different curve fitting models like polynomial, linear, spline, etc. In this letter, a special model of nonlinear curve fitting using rational functions is utilized in order to extract new features for classification and compression purposes. Therefore, it is important that $f(\lambda)$ has fewer parameters than the number of data samples.

Consider a function f and the two integers L ≥ 0 and M ≥ 0, the rational function approximant of order (L,M) for f is defined as:

$$[L/M]_f(\lambda) \triangleq \hat{f}(\lambda) = \frac{\sum_{j=0}^{L} c_{j+M+1}\lambda^j}{1+\sum_{j=1}^{M} c_j \lambda^j} \qquad (2)$$

Since, dissimilarities of curves yields differences in the coefficients of the approximants of the corresponding curves, it seems that these coefficients could be used as discriminating features for the curves. For example, in figure 1(a) two different families of curves are plotted, and the histogram of the coefficients of their rational function approximants are demonstrated in figure 1(b)-(f). In this curve fitting problem, we fitted a rational function with L=0 and M=4 to the both families of curves. Therefore, each curve can be expressed with its own five coefficients. As can be seen, the histograms of some coefficients completely separate the two families. This fact was our motivation for using the coefficients of the rational function fitted curve of the SRCs as discriminating features in HS data classification tasks.

3. The Proposed Feature Extraction Method

We introduced the spectral response curve (SRC) of each pixel of a hyperspectral image as the plot of its measured intensities in different wavelengths versus wavelengths or band numbers.

In other words, each SRC can be considered as a plot of a function $f(\lambda)$. As a matter of fact, we do not know the exact mathematical expression of $f(\lambda)$, but we have the amounts of its N consecutive points. We show that an approximation of $f(\lambda)$ in the form of a rational function with polynomial numerator and denominator can be developed through an LS method with uniform weights in (1). Then, we demonstrate that the coefficients of these polynomials can be applied as new features for maximum likelihood classifier with satisfactory results. Also these features can be used for reproducing the original data.

To provide equal condition for all hyperspectral images, the variable λ is considered in the form λ/N as normalized band number. The rational approximant of $f(\lambda/N)$ for the pixel located at (x,y) is given by:

$$\hat{f}_{(x,y)}\left(\frac{\lambda}{N}\right) = \frac{\sum_{j=0}^{L} c_{j+M+1}(\frac{\lambda}{N})^j}{1+\sum_{j=1}^{M} c_j(\frac{\lambda}{N})^j} \qquad (3)$$

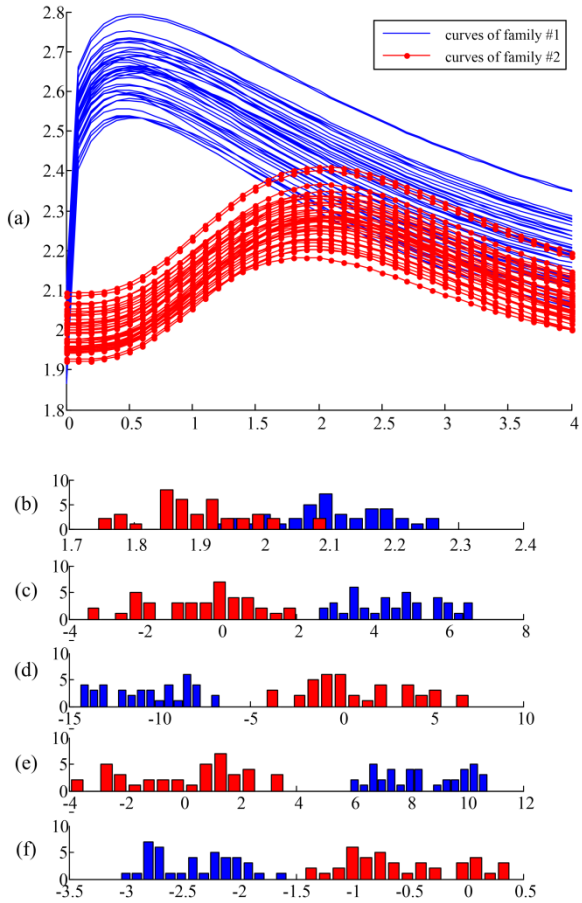


Fig. 1. Two families of curves; and the histograms of the coefficients of their corresponding Padé approximants, (a) Curves, blue (solids) and red (dotted) lines (b) numerator coefficient distribution, (c)-(f) denominator coefficients distribution

We want to determine coefficients vector $\boldsymbol{c}=[c_1\ c_2\ \ldots c_{M+L+1}]^t$ to minimize:

$$E = \frac{1}{N}\sum_{\lambda=1}^{N}\left(\hat{f}(\lambda/N) - f(\lambda/N)\right)^2 \qquad (4)$$

By computing partial derivatives of E with respect to the coefficients and setting them to zero, a system of nonlinear equations is obtained. A sufficient but not necessary condition for solving this system is to find **c** such that:

$$\sum_{j=0}^{L} c_{j+M+1}\left(\frac{\lambda}{N}\right)^j - f\left(\frac{\lambda}{N}\right)\sum_{j=1}^{M} c_j\left(\frac{\lambda}{N}\right)^j = f\left(\frac{\lambda}{N}\right); \qquad \lambda = 1,\ldots,N \qquad (5)$$

Now we have a system of linear equations with $M+L+1$ unknowns and N equations which can be rewritten in matrix form as:

$$A_{N\times(M+L+1)}\,c_{(M+L+1)\times 1} = B_{N\times 1} \qquad (6)$$

where,

$$\begin{cases} A_{N\times(M+L+1)} = [a_{\lambda j}] \\ a_{\lambda j} = \begin{cases} -f\left(\frac{\lambda}{N}\right).\left(\frac{\lambda}{N}\right)^j & j = 1,\ldots,M \\ \left(\frac{\lambda}{N}\right)^j & j = M+1,\ldots,M+L+1 \end{cases} \\ B = \left[f\left(\frac{1}{N}\right), f\left(\frac{2}{N}\right),\ldots,f\left(\frac{N}{N}\right)\right]^t \end{cases}$$

For feature reduction purposes, we want $M+L+1<<N$, therefore the matrix **A** is not square and the system of linear equations in (6), does not have a unique solution and sometimes may not have any solution. Instead, we use Moore-Penrose pseudo inverse of **A** and find **c** such that the norm of **Ac-B** be minimized.

The above procedure must be performed for all pixels of the hyperspectral data set and the vectors **c** replace the original data and new image cube is developed. Therefore the third dimension of data is changed to $M+L+1$, achieving to resize data for a rate of $N/(M+L+1)$. The procedure can be performed for all pixels, simultaneously.

This rational function curve fitting feature extraction (RFCF-FE) method is an unsupervised feature extraction method. The new features are applied to Maximum Likelihood (ML) classifiers and results are compared to PCA as a traditional unsupervised feature extraction method and LDA as a supervised feature extraction method. As it is demonstrated in next section, the RFCF-FE method results are more accurate than its competing methods for both urban and agricultural data sets. Also, it can be used as a coding algorithm for hyperspectral data compression.

4. Experimental Results

4.1 Hyperspectral Data Sets

The first data set used in our experiments is a mixed forest/agricultural 145×145 pixels image from Indian Pine Site (IPS) in Indiana [25]. It is captured by the Air-borne Visible/Infrared Imaging Spectrometer (AVIRIS). The spatial resolution of this data set is 20m. The picture contains 220 spectral bands in the wavelength range from 400 to 2500 nm with 10nm resolution. After removing twenty water absorption bands, N=200 bands were left. The class map of data contains 16 different land covers.

The other data set was gathered over the urban area of the University of Pavia (UP), by the Reflective Optics System Imaging Spectrometer (ROSIS) [25]. The image size is 610×340 pixels, the spatial resolution of 1.3m and the number of spectral bands is 115 in the wavelength range from 430 to 860 nm. Discarding some noisy bands yields N=103 bands, finally. Its ground truth map contains 9 different classes of land cover.

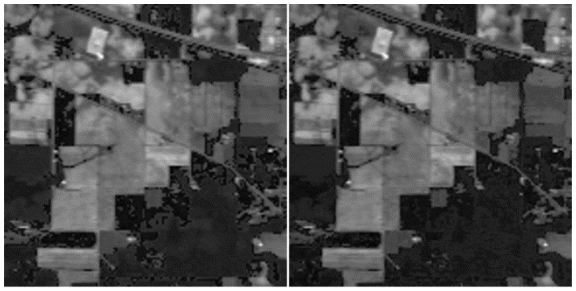
As an example, Images of band 18 of these data sets are demonstrated in Figures 2 (a) and 3 (a), respectively.

4.2 Results and Discussion

The proposed feature extraction method has been applied to both hyperspectral data sets IPS and UP, and the extracted features have been fed into an ML classifier. The classification results have been compared to those of PCA and LDA features, the two basic and traditional feature reduction approaches. For a given number of features, D, the parameter L has been changed in the range of 0 to D-1. Then M is selected regarding to the constraint: $M+L+1 = D$. For each value of D, the values

of L and M producing best results have been selected for comparison to PCA and LDA with the same dimensions. Since, 10 percent of the whole data volume with a minimum of 15 and a maximum of 50 samples per class is used for training the classifier, D is changed in the range $2 \leq D \leq 14$. Figures 2-(b) and 3-(b) show the images of 18th band of input data sets, reconstructed from extracted features by the RFCF-FE method (L=0, M=13 for IPS and L=12, M=1 for UP, as two typical amounts of M and L). Reconstruction process has been performed using (4) for λ=1, 2,…, N. Examples of real SRCs and their rational fitted curves of these two data sets are plotted in figure 4 where (L,M) are equal to (0,13) and (12,1) for IPS and UP, respectively.

The above discussions about Figures 2-4 show that the RFCF-FE method has high ability to preserve spatial characteristics of data as well as the spectral characteristics. Although, in a few points as it is illustrated in Figure 4, there may be a large difference between the original and the fitted curve, our results show that these exceptional points do not have severe impact on the classification performance. This is because: 1) these exceptional points in each SRC are rare, if any; 2) the locations of these points are not the same for different SRCs, i.e. this phenomenon does not destroy any specific band completely; 3) for each SRC, the coefficients of the corresponding fitted curve are used as the features, not the curve itself. As illustrated in Figure 4, the RFCF-FE method performs a smoothing operation in spectral domain. Moreover, it can be seen that by applying this method for IPS data some spatial domain smoothing has occurred without destroying edges of regions.



(a) (b)

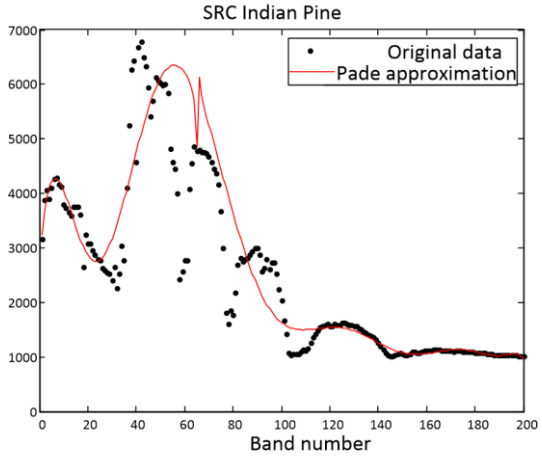
Fig. 2. Comparing the original and the reconstructed image of IPS (band 18) (a) Original data and (b) RFCF-FE method

(a) (b)

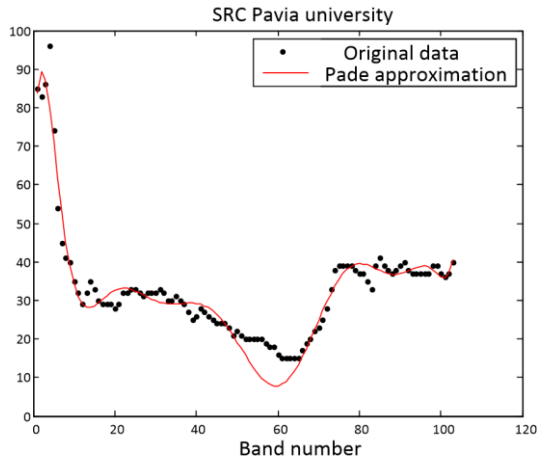
Fig. 3. Comparing the original and the reconstructed image of UP (band 18) (a) Original data, and (b) RFCF-FE method

Figures 5 and 6 demonstrate the accuracy assessment measures for IPS and UP data sets, respectively. These results are obtained by averaging the results of ten Monte Carlo runs (the standard deviations are illustrated by error bars). Also, Table 1 contains the comparison results in terms of parameter Z (see appendix A). The optimum values of L (and so M) parameters of the RFCF-FE method differ for different values of D and different iterations of the algorithm, but in the most cases, the best classification results have occurred when values of L are 0, 1, D-2, and D-1. Note that, despite PCA and the RFCF-FE method, maximum number of extracted features in LDA method is equal to N_c-1, where N_c is the number of classes.

The superiority of the RFCF-FE method compared to PCA and LDA methods is apparent from this table and Figures 5 and 6. As it can be seen, all measures (average accuracy, average validity, overall accuracy, kappa statistics, and Z) have been dramatically improved by RFCF-FE algorithm in comparison to both PCA and LDA. In terms of the first four accuracy measures, this improvement in IPS is more than in UP because of its larger ground pixels and agricultural nature with relatively large uniform areas and less details. Whereas, in terms of Z the improvement in UP is much more obvious. Note that average accuracy, average validity, overall accuracy, and kappa statistics focus on pure average accuracies despite Z parameter which focuses on disagreements between two algorithms and determines the superior one. Therefore, our method excels PCA and LDA for agricultural scenes as well as urban ones.



(a)



(b)

Fig. 4. SRC of pixel (20,20) and its approximation (a) IPS data and (b) UP

Table 2 contains PSNR (see appendix B) values for reconstructed images from PCA and the proposed method for IPS and UP, respectively when D varies from 3 until 15. PSNR for the proposed method corresponds to the values of L and M that yields to the best result. These optimum values of L and M have been shown in these tables. The superiority of the proposed method with respect to PCA based compressing method is apparent from these tables. As it is demonstrated in Table 2, in most cases for UP data set, the best result happens when M=0. It implies that McLaurin series has better performance in these cases.

Another advantage of the proposed method is that unlike the competitive methods, it is applied pixel by pixel and does not need to transform the data to the projection space as a whole. Therefore, real-time or parallel implementation of the algorithm is possible. However, the complexity of the method might be higher than the competitive algorithms. Moreover, since the proposed transform is invertible, it could be used as a compression algorithm for HS data.

Table 1. Parameter Z (mean± standard deviation of 10 MonteCarlo runs) of ML classifier for RFCF features vs. PCA and LDA

No. of Features	IPS data set		UP data set	
	RFCF-FE vs. PCA	RFCF-FE vs. LDA	RFCF-FE vs. PCA	RFCF-FE vs. LDA
2	5.87 ± 4.96	5.50 ± 1.99	0.10±0.30	0.07±0.23
3	7.60 ± 0.94	8.47 ± 1.36	15.54 ± 6.07	16.15±5.91
4	2.87 ± 1.12	5.27 ± 2.30	7.15±4.54	8.38±3.53
5	0.37 ± 0.40	1.86 ± 1.70	12.51±5.24	13.55±5.19
6	5.32 ± 4.33	5.56 ± 3.76	13.91±4.86	18.89±3.94
7	2.40± 1.79	2.50 ± 2.63	19.41±4.24	22.90±3.44
8	3.19 ± 3.12	4.11 ± 4.05	18.40±5.25	20.98±5.40
9	1.04 ± 1.25	3.07 ± 3.33	24.03±6.10	-
10	0.07 ± 0.21	0.27 ±0.49	23.82±4.80	-
11	0.26 ± 0.61	1.27 ± 1.07	21.62±4.70	-
12	3.67 ± 1.07	4.87 ± 1.91	23.50±6.23	-
13	0.67±0.81	1.36±1.50	36.25±6.90	-
14	4.12±1.74	4.04±1.87	36.21±6.57	-

Table 2. Comparing PSNR of the proposed method and inverse PCA for IPS and UP data sets in terms of compression rate (N/D)

Rate *(N/D)*	IPS			Rate *(N/D)*	UP		
	Best *(L,M)*	PSNR	IPCA PSNR		Best *(L,M)*	PSNR	IPCA PSNR
200/3	(0,2)	34.95	24.39	103/3	(2,0)	31.85	29.76
200/4	(0,3)	27.42	25.74	103/4	(3,0)	20.25	16.44
200/5	(0,4)	58.70	24.20	103/5	(2,2)	32.92	23.54
200/6	(0,5)	29.63	23.28	103/6	(5,0)	21.35	16.17
200/7	(4,2)	24.17	24.46	103/7	(5,1)	25.29	19.27
200/8	(0,7)	22.78	25.13	103/8	(2,5)	24.06	19.52
200/9	(6,2)	55.11	25.21	103/9	(1,7)	43.77	22.52
200/10	(2,7)	34.38	25.90	103/10	(2,7)	38.98	21.77
200/11	(6,4)	28.46	27.03	103/11	(6,4)	68.19	22.74
200/12	(8,3)	28.17	28.99	103/12	(11,0)	34.11	23.65
200/13	(5,7)	31.47	29.43	103/13	(12,0)	27.86	22.91
200/14	(2,11)	41.01	31.47	103/14	(13,0)	26.05	24.08
200/15	(10,4)	42.38	30.14	103/15	(14,0)	35.47	23.98

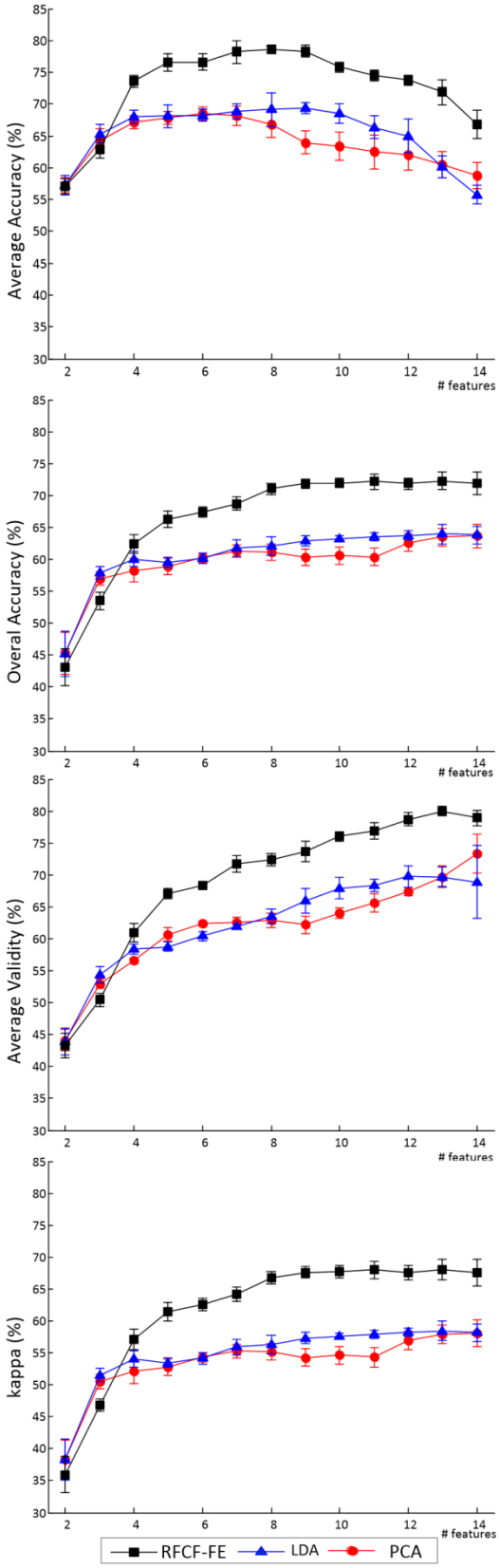


Fig. 5. Comparison of the RFCF-FE method with PCA and LDA for IPS data

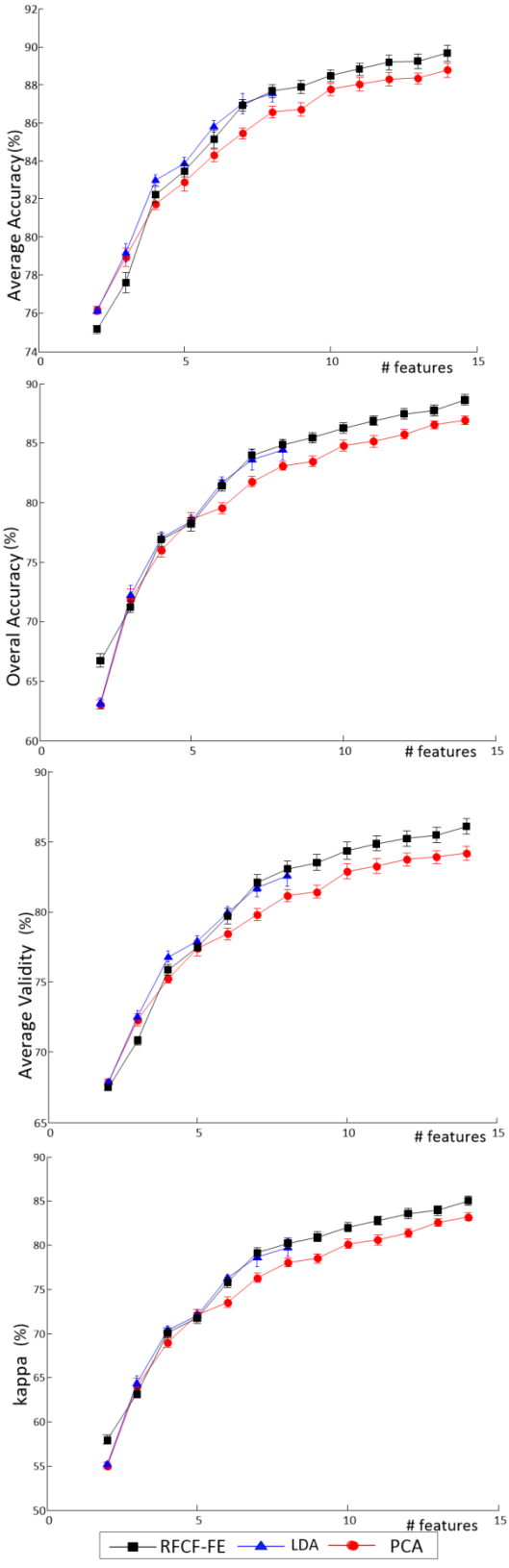


Fig. 6. Comparison of the RFCF-FE method with PCA and LDA for UP data

5. Conclusions

A new feature extraction method for hyperspectral data was proposed based on rational function curve fitting. The main motivation for using curve fitting approach for hyperspectral data feature extraction is the utilization of the information that exists in the sequence of original features (ordinance of reflectance coefficients in SRC) that are neglected by competing methods. The coefficients of SRCs approximants are calculated through analytical operations. These extracted features are then fed into an ML classifier. The classification performance was compared to PCA and LDA as two classic and traditional F.E. methods. The results show the superiority of the proposed method. Also, it has been shown that this technique has satisfactory results for signal visualization and signal representation, and can be considered as a good coding algorithm for lossy compression of HS data. The proposed method is applied pixel by pixel and does not need to transform whole data to a new space simultaneously. Therefore, it can be applied to all pixels in a parallel procedure. In addition, this method is a novel approach which can be used as a powerful base for developing more efficient feature extraction methods.

6. Acknowledgement

This research is done by support of Iran communication research centre under contract number 18133/500 T by Identification code: 90-01-03. The authors gratefully acknowledge that organization for its support.

Appendices

A. McNemar's test:

In a classification assessment procedure, the statistical significance of differences can be computed using McNemar's test. This test is based upon the standardized normal test statistic $Z = (n_{12} - n_{21})/\sqrt{n_{12} + n_{21}}$ where n_{12} is the number of samples classified correctly by classifier 1, but wrongly by classifier 2. At the 5% level of significance, if the difference in accuracy between classifiers 1 and 2 is statistically significant. When classifier 1 is more accurate than classifier 2, *Z* is positive or vice versa [26].

B. Average Classification Accuracy (AA)

Classification accuracy for each individual class is calculated as ACC(c) = nc/Nc, in which, nc is the number of the pixels of class *c* correctly classified, and N_c is the number of test pixels in that class. Average classification accuracy (AA) is then defined as follows:

$$\mathrm{AA} = \frac{1}{\mathrm{C}}\sum_{\mathrm{c}=1}^{\mathrm{C}} \mathrm{ACC(c)} \qquad (7)$$

where, C is the number of all classes.

C. Average Classification Validity (AV)

Classification validity (or reliability) for each individual class is calculated as $\mathrm{VAL(c)} = n_c/m_c$ in which, n_c is the number of the pixels of class *c* correctly classified, and m_c is the number of all the pixels labelled as class *c* in output class map. Average classification validity (AV) is then defined as:

$$\mathrm{AV} = \frac{1}{\mathrm{C}}\sum_{\mathrm{c}=1}^{\mathrm{C}} \mathrm{VAL(c)} \qquad (8)$$

D. Overall Classification Accuracy (OA)

This measure is similar to average classification accuracy, except that classes are not considered individually, but as a whole:

$$\mathrm{OA} = \frac{\mathrm{n}}{\mathrm{N}} \qquad (9)$$

in which, n is the number of all the pixels correctly classified, regardless of the class label, and N is the number of all pixels in the test set.

E. Kappa coefficient

This coefficient is the overall accuracy (OA) corrected by the amount of agreement that is expected due to chance:

$$kappa = \frac{OA - P_e}{1 - P_e} \; ; in\ which\ \ P_e = \frac{\sum_{c=1}^{C} m_c \,.\, N_c}{N^2} \qquad (10)$$

in which, OA is the overall accuracy mentioned above, m_c is the number of all the pixels labelled as class *c* in the output class map.

F. PSNR:

PSNR is defined as: 10 log*(S/N)*, where *S* is the energy of original signal and *N* is the energy of difference between original and decompressed signal. Also, compression rate is defined as the ratio of the original to the compressed signal volume.

References

[1] D. Landgrebe, "Hyperspectral image data analysis as a High Dimensional Signal Processing Problem," *Signal Processing Magazine, IEEE*, vol. 19, pp. 17-28, 2002.

[2] B. M. Shahshahani and D. A. Landgrebe, "The effect of unlabeled samples in reducing the small sample size problem and mitigating the Hughes phenomenon," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 32, pp. 1087-1095, 1994.

[3] M. Marconcini, G. Camps-Valls, and L. Bruzzone, "A Composite Semisupervised SVM for Classification of Hyperspectral Images," *Geoscience and Remote Sensing Letters, IEEE*, vol. 6, pp. 234-238, 2009.

[4] H. Ghassemian and D. A. Landgrebe, "Object-oriented feature extraction method for image data compaction," *Control Systems Magazine, IEEE*, vol. 8, pp. 42-48, 1988.

[5] G. Camps-Valls, N. Shervashidze, and K. M. Borgwardt, "Spatio-Spectral Remote Sensing Image Classification With Graph Kernels," *Geoscience and Remote Sensing Letters, IEEE*, vol. 7, pp. 741-745, 2010.
[6] F. Mirzapour and H. Ghassemian, "Improving hyperspectral image classification by combining spectral, texture, and shape features," *International Journal of Remote Sensing*, vol. 36, pp. 1070-1096, 2015/02/16 2015.
[7] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 42, pp. 1778-1790, 2004.
[8] J. Xiuping, K. Bor-Chen, and M. M. Crawford, "Feature Mining for Hyperspectral Image Classification," *Proceedings of the IEEE*, vol. 101, pp. 676-697, 2013.
[9] G. Lianru, L. Jun, M. Khodadadzadeh, A. Plaza, Z. Bing, H. Zhijian, et al., "Subspace-Based Support Vector Machines for Hyperspectral Image Classification," *Geoscience and Remote Sensing Letters, IEEE*, vol. 12, pp. 349-353, 2015.
[10] G. H. Halldorsson, J. A. Benediktsson, and J. R. Sveinsson, "Source based feature extraction for support vector machines in hyperspectral classification," in *Geoscience and Remote Sensing Symposium, 2004. IGARSS '04. Proceedings. 2004 IEEE International*, 2004, p. 539.
[11] R. Pu and P. Gong, "Wavelet transform applied to EO-1 hyperspectral data for forest LAI and crown closure mapping," *Remote Sensing of Environment*, vol. 91, pp. 212-224, 5/30/ 2004.
[12] H. Xiaofei, C. Deng, and H. Jiawei, "Learning a Maximum Margin Subspace for Image Retrieval," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 20, pp. 189-201, 2008.
[13] D. Qian, "Modified Fisher's Linear Discriminant Analysis for Hyperspectral Imagery," *Geoscience and Remote Sensing Letters, IEEE*, vol. 4, pp. 503-507, 2007.
[14] W. Jing and I. C. Chein, "Independent component analysis-based dimensionality reduction with applications in hyperspectral image analysis," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 44, pp. 1586-1600, 2006.
[15] G. Yanfeng, L. Ying, and Z. Ye, "A Selective KPCA Algorithm Based on High-Order Statistics for Anomaly Detection in Hyperspectral Imagery," *Geoscience and Remote Sensing Letters, IEEE*, vol. 5, pp. 43-47, 2008.
[16] S. Mika, G. Ratsch, J. Weston, B. Scholkopf, and K. Muller, "Fisher discriminant analysis with kernels," in *Neural Networks for Signal Processing IX, 1999. Proceedings of the 1999 IEEE Signal Processing Society Workshop.*, 1999, pp. 41-48.
[17] G. Baudat and F. Anouar, "Generalized discriminant analysis using a kernel approach," *Neural computation*, vol. 12, pp. 2385-2404, 2000.
[18] M. Pal and P. M. Mather, "An assessment of the effectiveness of decision tree methods for land cover classification," *Remote Sensing of Environment*, vol. 86, pp. 554-565, 8/30/ 2003.
[19] A. Hosseini and H. Ghassemian, "Classification of hyperspectral and multispectral images by using fractal dimension of spectral response curve," in *Electrical Engineering (ICEE), 2012 20th Iranian Conference on*, 2012, pp. 1452-1457.
[20] S. A. Hosseini and H. Ghassemian, "A new hyperspectral image classification approach using fractal dimension of spectral response curve," in *Electrical Engineering (ICEE), 2013 21st Iranian Conference on*, 2013, pp. 1-6.
[21] M. Unser, A. Aldroubi, and M. Eden, "B-spline signal processing. I. Theory," *Signal Processing, IEEE Transactions on*, vol. 41, pp. 821-833, 1993.
[22] D. A. Wagenaar and S. M. Potter, "Real-time multi-channel stimulus artifact suppression by local curve fitting," *Journal of Neuroscience Methods*, vol. 120, pp. 113-120, 10/30/ 2002.
[23] L. Fang and D. C. Gossard, "Multidimensional curve fitting to unorganized data points by nonlinear minimization," *Computer-Aided Design*, vol. 27, pp. 48-58, 1// 1995.
[24] H. Motulsky and A. Christopoulos, *Fitting models to biological data using linear and nonlinear regression: a practical guide to curve fitting*: Oxford University Press, 2004.
[25] M. Kamandar and H. Ghassemian, "Linear Feature Extraction for Hyperspectral Images Based on Information Theoretic Learning," *Geoscience and Remote Sensing Letters, IEEE*, vol. 10, pp. 702-706, 2013.
[26] G. M. Foody, "Thematic Map Comparison: Evaluating the Statistical Significance of Differences in Classification Accuracy," P*hotogrammetric Engineering & Remote Sensing*, vol. 70, pp. 627-633, 05/01/ 2004.

Seyed Abolfazl Hosseini received the B.Sc. degree in 1993 and M.Sc. degree in 1997 in Electrical Engineering from Sharif University of technology and K.N.Toosi University of technology, Tehran, Iran, respectively. He is currently working toward the Ph.D. degree in the Faculty of Electrical and Computer Engineering at the Tarbiat Modares University, Tehran, Iran. His research interests include the hyperspectral image analysis, feature extraction and pattern recognition in remote sensing applications.

Hassan Ghassemian received his B.Sc. degree from Tehran College of Telecommunication in 1980 and the M.Sc. and Ph.D. degree from Purdue University, West Lafayette, USA in 1984 and 1988 respectively. Since 1988, he has been with Faculty of Computer and Electrical Engineering at Tarbiat Modares University in Tehran, Iran, where he is a Professor of Electrical and Computer Engineering. Dr. Ghassemian has published more than 370 technical papers in peer-reviewed journals and conference proceedings. His current research interests focus on Multi-Source Signal/Image Processing, Information Analysis and Remote Sensing.