Feature reduction of hyperspectral images: Discriminant analysis and the first principal component

M. Imani and H. Ghassemian[*]

Faculty of Electrical & Computer Engineering, Tarbiat Modares University, Tehran, Iran

Abstract

When the number of training samples is limited, feature reduction plays an important role in classification of hyperspectral images. In this paper, we propose a supervised feature extraction method based on discriminant analysis (DA) which uses the first principal component (PC1) to weight the scatter matrices. The proposed method, called DA-PC1, copes with the small sample size problem and has not the limitation of linear discriminant analysis (LDA) in the number of extracted features. In DA-PC1, the dominant structure of distribution is preserved by PC1 and the class separability is increased by DA. The experimental results show the good performance of DA-PC1 compared to some state-of-the-art feature extraction methods.

Keywords: *Discriminant Analysis, Principal Component, Feature Reduction, Hyperspectral, Classification.*

1. Introduction

Due to the recent advances of remote sensing instruments, hyperspectral imaging has become a fast growing technique in the field of remote sensing [1]. Hyperspectral imaging sensors with acquiring a large number of spectral bands allows us to better distinguish many subtle objects and materials [2]. An important application of hyperspectral imaging is image classification [3-6]. However, as the inputs of hyperspectral datasets are high-dimensional vectors whose coordinates are highly correlated, the direct use of classical models for hyperspectral image classification faces several difficulties particularly when the number of available training samples is limited. With a fixed number of training samples, hyperspectral image classification accuracy can first increase as the dimensionality of data increases, but decays with the dimensionality higher than some optimum value. In other words, Hughes phenomenon occurs [7]. One of the main approaches to mitigate this problem is dimensionality reduction [8-11].

Feature reduction can be done with feature selection or feature extraction. In feature selection approaches, just an appropriate subset of original features is selected usually using a discrimination criterion and a search algorithm [12-18]. Thus, the physical meaning of data is preserved using the feature selection. However, in the feature extraction method, a linear or nonlinear transformation is applied to the original features to extract some new features [19-26]. Depending on the use of labeled samples for training feature extraction, theses techniques are divided into supervised ones, which use the class label information, and unsupervised ones, which do not use the class label information for training.

Principal component analysis (PCA) and linear discriminant analysis (LDA) are the most widely used unsupervised and supervised linear feature extraction methods, respectively [27]. PCA finds the principal components in accordance with the maximum variance of a data matrix. Thus, after such a transformation, the dominant structure of the distribution can be well preserved in the reduced subspace. The generated principal components are linear combination of the original features and are uncorrelated. PCA searches directions with a large variance in the data and subsequently projects data onto it. The position of data in the reduced feature space may be inappropriate to distinguish between classes to

have a good classification. LDA utilizes the label information to infer class separability. LDA seeks projection directions on which the ratio of the between-class scatter to within-class scatter is maximized. Some difficulties with LDA method are as follows. When the number of training samples is limited, the accurate estimate of scatter matrices may not be obtained and the within-class scatter matrix becomes singular. Thus LDA has no reasonable performance in small sample size situation. Moreover, LDA can extract maximum $c-1$ features (c is the number of classes) which is not always sufficient for representing the original data.

As the generalized discriminant analysis (GDA) provides a mapping of input vectors into a high dimensional feature space, it can deal with nonlinear discriminant analysis using kernel function operator [28]. Using a different kernel, one can cover a wide class of nonlinearities. GDA, which is the kernelized version of LDA, can extract maximum $c-1$ features. Nonparametric weighted feature extraction (NWFE) has been proposed for improving LDA [29]. To put different weights on samples to compute the weighted means, defining new nonparametric between-class and within-class scatter matrices to obtain more than $c-1$ features is the main idea of NWFE. In order to alleviate the negative influence of outliers in class-mean based methods, authors in [30] have proposed a novel linear dimensionality reduction technique called median–mean line based discriminant analysis (MMLDA) method. They rectify to some extent the position of the class-mean caused by outliers by introducing the median–mean line as an adaptive class-prototype.

In this paper, we propose a supervised feature extraction method based on discriminant analysis (DA). The proposed method uses the first principal component (PC1) to weight scatter matrices. So, in addition to class discrimination information contained in the Fisher criterion (maximizing the between-class scatter and minimizing the within-class scatter), the proposed method can use the data representation and reconstruction information to preserve the main structure of original data in the reduced subspace. Moreover, the non-parametric form of scatter matrices and the use of regularization method help to extraction of more than $c-1$ features and also to solve the singularity problem.

We introduce the proposed method in section 2, called DA-PC1, with more details. Then, in section 3, the extensive experiments show that the proposed method outperforms popular feature

extraction methods in terms of classification accuracy. Finally, the conclusions are discussed in section 4.

2. DA-PC1

The proposed feature extraction method, DA-PC1, uses the discriminant analysis to increase the separability between classes. DA-PC1 maximizes the between-class scatter matrix and minimizes the within-class scatter matrix. It defines the weighted non-parametric scatter matrices to provide three main advantages: 1- DA-PC1 copes with the singularity problem of within-class scatter matrix in the small sample size situation. 2- It can extract more than $c-1$ features where c is the number of classes. 3- In addition to class discrimination information, it uses the reconstruction information contained in the first principal component for weighting the scatter matrices. In the first step, we compute the first principal component (PC1) of data. For reaching this purpose, we estimate the covariance matrix of data as follows:

$$\Sigma_x = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \overline{x})\,(x_i - \overline{x})^T \tag{1}$$

where, $x_i \in \mathcal{R}^d$ ($i = 1,2,...,N$) is the ith pixel of hyperspectral image, d is the number of spectral bands, N is the total number of samples (pixels) and $\overline{x}$ is the total mean of data that is given by:

$$\overline{x} = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{2}$$

The PC1 is obtained by using the eigenvector v_1 correspondence with the largest eigenvalue of Σ_x. Then, we have:

$$PC1(x_i) = v_1^T x_i \,;\, (i = 1,2,...,N) \tag{3}$$

The between-class scatter matrix ($\mathbf{S}_b$) and the within-class scatter matrix ($\mathbf{S}_w$) are calculated as follows:

$$\mathbf{S}_w = \sum_{j=1}^{n}\sum_{\substack{i=1 \\ l_i = l_j}}^{n} w_{ij}\big(x_{ti} - x_{tj}\big)\big(x_{ti} - x_{tj}\big)^T \tag{4}$$

$$\mathbf{S}_b = \sum_{j=1}^{n}\sum_{\substack{i=1 \\ l_i \neq l_j}}^{n} w_{ij}\big(x_{ti} - x_{tj}\big)\big(x_{ti} - x_{tj}\big)^T \tag{5}$$

where, x_{ti} ($i = 1, 2, ..., n$) is the i th training sample, n is the total number of training samples and $l_i \in \{1, 2, ..., c\}$ is the class label of sample x_{ti}, and c is the number of classes.

The closer the principal components of two samples x_{ti} and x_{tj} are, the larger weight w_{ij} will be. Thus, the weight w_{ij} ($i = 1, ..., n; j = 1, ..., n$) is calculated as follows:

$$w_{ij} = \frac{1}{\left|PC1(x_{ti})-PC1(x_{tj})\right|^2+1} \quad (6)$$

The number one is added to the denominator because w_{ij} should not be infinite. In above equation, we have:

$$PC1(x_{ti}) = v_1^T x_{ti} \; ; \; (i = 1, 2, \dots, n) \quad (7)$$

To degrade the singularity problem and thus, to increase the classification accuracy, we regularize the matrix S_w as follows:

$$S_w = 0.5S_w + 0.5diag(S_w) \quad (8)$$

Because of non-parametric form of S_w and also with the regularization of it, DA-PC1 copes with the singularity problem in small sample size situation. Because of non-parametric form of S_b, DA-PC1 can extract more than $c-1$ features. DA-PC1 uses both the information contained in the DA and PC1. DA-PC1 increases the class separability using DA. Moreover, the PC1, which is in accordance with the maximum variance of data matrix, can preserve the dominant structure of distribution in the reduced subspace after transformation.

3. Experiments and results

The performance of DA-PC1 is compared with LDA, NWFE, GDA, MMLDA, and PCA. To assess the performance of classification, we use the accuracy and reliability of classes, the average accuracy, the average reliability, kappa coefficient [31] and also the McNemar test results [32]. The definitions of these measures are represented below.

The accuracy (Acc.) and reliability (Rel.) for each class are defined as $Acc = \gamma/\alpha$ and $Rel = \gamma/\beta$ respectively where γ is the number of testing samples that are correctly classified, α denotes the total testing samples of class and β is the total samples which are labeled as this class. The kappa coefficient is defined as follows:

$$Kappa = \frac{N\sum_{k=1}^{c} t_{kk} - \sum_{k=1}^{c} t_{k+}t_{+k}}{N^2 - \sum_{k=1}^{c} t_{k+}t_{+k}} \quad (9)$$

where, N and c denote the number of testing samples and the number of classes, respectively. t_{kk} is the number of samples correctly classified in class k, t_{k+} is the number of testing samples labeled as class k, and t_{+k} is the number of samples predicted as belonging to class k. The McNemar test is used to assess the statistical significance of differences in classification results. The parameter Z_{12} in McNemar test is defined as follows:

$$Z_{12} = \frac{f_{12}-f_{21}}{\sqrt{f_{12}+f_{21}}} \quad (10)$$

where, f_{12} is the number of samples which are labeled correctly by classifier 1 and incorrectly by classifier 2. The difference in the accuracy between two classifiers is said to be statistically significant if $|Z_{12}| > 1.96$. If classifier 1 is more accurate than classifier 2, we have $Z_{12} > 0$ and otherwise $Z_{12} < 0$.

Kennedy Space Center (KSC) was acquired by the NASA Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) instrument over the KSC, Florida, on March 23, 1996. It contains 13 ground-truth classes and 512×614 pixels. This hyperspectral image has 224 spectral channels which after discarding water absorption and noisy bands, 176 bands are retained. Indian Pines dataset was acquired by the AVIRIS sensor in 1992. This image scene contains 145×145 pixels and 16 classes which 10 classes of it are chosen for our experiments. The Indian dataset has 220 spectral bands, where 20 bands were discarded because of the atmospheric affection. The Pavia University dataset was provided by the Reflective Optics System Imaging Spectrometer (ROSIS) with a spatial resolution of 1.3 m per pixel. The number of spectral channels in the original recorded image is 115 (with a spectral range from 0.43 to 0.86 μm) which after the removal of the noisy bands, 103 spectral bands are selected. This urban image contains nine classes and 610×340 pixels.

We used three different classifiers for classification of reduced datasets: support vector machine (SVM), maximum likelihood (ML), and nearest neighbor (NN). The Gaussian distribution is assumed for datasets in ML classifier. We use the Radial basis function (RBF) kernel and the one-against-one multiclass classification algorithm for SVM classification. We test the penalty parameter C of SVM between [10– 1000] with a step size increment of 20 and the γ parameter of the RBF kernel between [0.1– 2] with a step size increment of 0.1. The best values of free parameters are obtained using a 5-fold cross validation approach.

The training samples are chosen randomly from entire datasets and the remaining samples are used for testing. Each experiment is repeated 10 times, with different random training samples in each time, and the average results are reported.

The average classification accuracies versus the number of extracted features are shown in figure 1 for Indian dataset with a) SVM classifier, 10 training samples, b) ML classifier, 10 training samples, c) NN classifier, 10 training samples, d)

SVM classifier, 15 training samples, e) ML classifier, 15 training samples, f) NN classifier, 15 training samples, g) SVM classifier, 30 training samples, h) ML classifier, 30 training samples, i) NN classifier, 30 training samples, j) SVM classifier, 60 training samples, k) ML classifier, 60 training samples, l) NN classifier, 60 training samples. In most cases, the better performance of DA-PC1 compared to other feature extraction methods can be seen.

Figure 1 shows that the efficiency of LDA using just 10 and 15 training samples is very weak because of singularity of within-class scatter matrix. Moreover, LDA and GDA can extract maximum $c - 1 = 9$ features which are insufficient in some cases for accurate classification of data.

The classification accuracies of reduced data in different number of extracted features with using 15 training samples per class for KSC dataset obtained by a) SVM classifier, b) ML classifier, and c) NN classifier are shown in figure 2. The accuracy and reliability of classes obtained by 15 training samples and SVM classifier are represented for Indian (with 6 extracted features) and KSC (with 8 extracted features) datasets in table 1 and table 2, respectively.

The McNemar test results and the ground truth map (GTM) and the classification maps of theses cases are shown in table 3, figures 3, and 4, respectively.

In table 3, Z_{rc} denotes each case of table where r is the row and c is the column. The highest classification accuracies are represented in table 4. The numbers in the parentheses are the number of features which achieve the highest average classification accuracy in each method. In most cases, DA-PC1 obtains the maximum classification accuracy compared to other methods.

Just in the following cases, other feature extraction methods obtain more classification accuracy compared to the proposed method. In Pavia University dataset, with using ML classifier, the best classification accuracy is obtained by GDA, MMLDA, and PCA methods and with using NN classifier, MMLDA obtains the best result for this dataset.

Moreover, in KSC dataset, with using ML classifier, the maximum classification accuracy is obtained by GDA. In figure 5, we compare the performance of DA-PC1 with PCA and LDA in a fixed number of extracted features with varying the number of training samples from 5 to 130

samples per class, for Indian dataset with a) SVM classifier and 6 extracted features, b) ML classifier and 5 extracted features, c) NN classifier and 9 extracted features. The following points can be concluded from the results of this experiment:

1) When the training set is small, PCA works better than LDA and when the training set is large, LDA works better than PCA.
2) When the number of training samples is limited, DA-PC1 is superior to both the PCA and LDA and when the high number of training samples is available, with using SVM and NN classifiers, LDA outperforms DA-PC1. However, in this case, DA-PC1 has yet reasonable performance.
3) With using ML classifier, the performance of DA-PC1 is better than PCA and LDA in both cases of small and large training set.

In general, when we use the parametric classifiers such as ML, which need to calculate the mean vectors and covariance matrices of classes and have more sensitivity to the training set, the use of DA-PC1 is preferable compared to PCA and LDA whether the training set size is small or large. But, when we use the non-parametric classifiers such as SVM and NN, which have the less sensitivity to the number of training samples, DA-PC1 is preferable using small training set and LDA is preferable with large training set.

4. Conclusion

In this paper, the DA-PC1 method is proposed for feature extraction of hyperspectral images. In DA-PC1, the first principal components of training samples are used for weighting of scatter matrices in the DA. Thus, in addition to increasing the class separability, the signal representation in the reduced subspace may be improved.

The experimental results show that with using parametric classifiers such as ML, the use of DA-PC1 is superior to PCA and LDA whether the training set size is small or large. Also, with using non-parametric classifiers such as SVM and NN, the use of DA-PC1 is superior to PCA and LDA with a small training sample size. Moreover, the comparison of DA-PC1 with the state-of-the-art feature extraction methods such as NWFE, GDA, and MMLDA shows the better performance of DA-PC1 for feature reduction and classification of hyperspectral images particularly in small sample size situation.

Figure 1. The average classification accuracies versus the number of extracted features for Indian dataset with a) SVM classifier, 10 training samples, b) ML classifier, 10 training samples, c) NN classifier, 10 training samples, d) SVM classifier, 15 training samples, e) ML classifier, 15 training samples, f) NN classifier, 15 training samples, g) SVM classifier, 30 training samples, h) ML classifier, 30 training samples, i) NN classifier, 30 training samples, j) SVM classifier, 60 training samples, k) ML classifier, 60 training samples, l) NN classifier, 60 training samples.
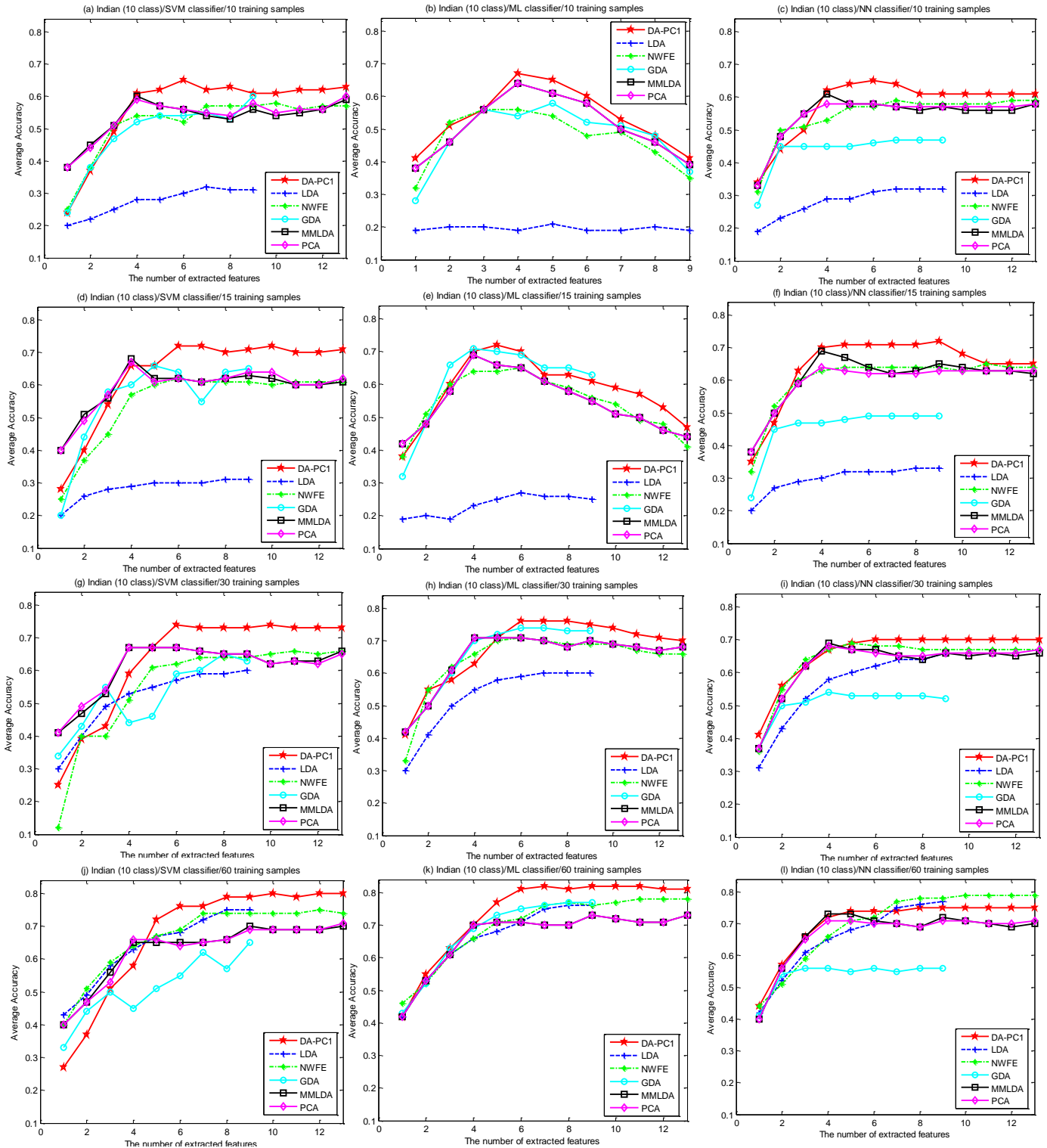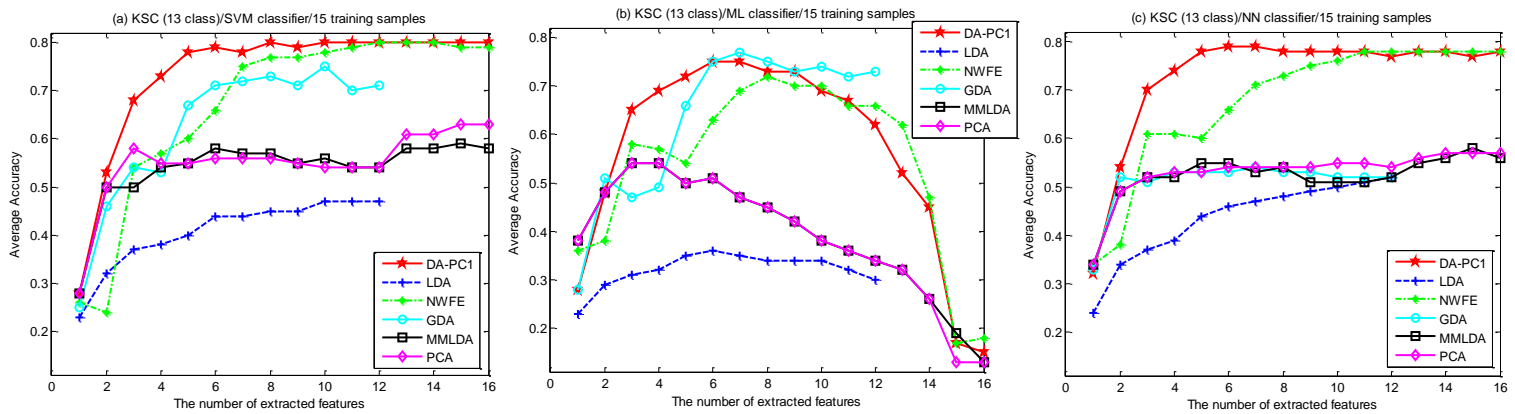
Figure 2. The average classification accuracies versus the number of extracted features for KSC dataset using 15 training samples obtained by a) SVM classifier, b) ML classifier, and c) NN classifier.

Table 1. The accuracy and reliability of classes of Indian dataset obtained by SVM classifier, 15 training samples and 6 extracted features.

	class		DA-PC1		LDA		NWFE		GDA		MMLDA		PCA	
No	Name of class	# samples	Acc.	Rel.	Acc.	Rel.	Acc.	Rel.	Acc.	Rel.	Acc.	Rel.	Acc.	Rel.
1	Corn-no till	1434	0.94	0.36	0.81	0.32	0.91	0.29	0.91	0.30	0.91	0.30	0.93	0.30
2	Corn-min till	834	0.77	0.53	0.21	0.37	0.43	0.41	0.60	0.46	0.46	0.39	0.45	0.41
3	Grass/pasture	497	0.68	0.44	0.10	0.25	0.57	0.35	0.53	0.34	0.35	0.22	0.39	0.26
4	Grass/trees	747	0.76	0.41	0.42	0.05	0.53	0.47	0.53	0.40	0.61	0.49	0.60	0.53
5	Hay-windrowed	489	0.79	0.62	0.42	0.47	0.74	0.57	0.45	0.65	0.79	0.45	0.78	0.45
6	Soybeans-no till	968	0.66	0.87	0.34	0.49	0.79	0.83	0.83	0.79	0.58	0.79	0.60	0.78
7	Soybeans-min till	2468	1.00	0.54	0.92	0.13	1.00	0.51	0.35	0.12	1.00	0.32	1.00	0.42
8	Soybeans-clean till	614	0.78	1.00	0.45	0.99	0.78	0.99	0.73	0.95	0.72	0.99	0.73	1.00
9	Woods	1294	0.95	0.20	1.00	0.02	0.95	0.13	0.95	0.50	0.95	0.36	1.00	0.42
10	Bldg-Grass-Tree-Drives	380	0.53	0.59	0.33	0.20	0.58	0.39	0.63	0.47	0.55	0.51	0.55	0.50
	Average Acc. and Average Rel.		**0.72**	**0.68**	**0.30**	**0.31**	**0.62**	**0.59**	**0.64**	**0.62**	**0.62**	**0.60**	**0.62**	**0.60**
	Kappa coefficient		**0.60**		**0.19**		**0.51**		**0.56**		**0.54**		**0.54**	

Table 2. The accuracy and reliability of classes of KSC dataset obtained by SVM classifier, 15 training samples and 8 extracted features.

	class		DA-PC1		LDA		NWFE		GDA		MMLDA		PCA	
No	Name of class	# samples	Acc.	Rel.	Acc.	Rel.	Acc.	Rel.	Acc.	Rel.	Acc.	Rel.	Acc.	Rel.
1	Scrub	761	0.80	0.95	0.50	0.59	0.67	0.96	0.77	0.91	0.72	0.93	0.68	0.92
2	Willow swamp	243	0.87	0.86	0.42	0.54	0.88	0.91	0.84	0.84	0.68	0.46	0.67	0.48
3	Cabbage palm hammock	256	0.88	0.80	0.29	0.26	0.82	0.81	0.87	0.76	0.72	0.74	0.59	0.75
4	Cabbage palm/oak hammock	252	0.56	0.52	0.11	0.23	0.70	0.44	0.50	0.32	0.25	0.30	0.30	0.31
5	Slash pine	161	0.63	0.43	0.24	0.25	0.42	0.42	0.30	0.30	0.40	0.30	0.43	0.26
6	Oak/broadleaf hammock	229	0.59	0.52	0.54	0.19	0.45	0.32	0.20	0.26	0.38	0.29	0.35	0.28
7	Hardwood swamp	105	0.93	0.64	0.51	0.27	0.90	0.75	0.85	0.76	0.48	0.41	0.49	0.46
8	Graminoid marsh	431	0.67	0.79	0.55	0.47	0.83	0.69	0.82	0.75	0.41	0.35	0.28	0.30
9	Spartina marsh	520	0.93	0.73	0.20	0.43	0.82	0.82	0.91	0.77	0.79	0.69	0.85	0.62
10	Cattail marsh	404	0.82	0.97	0.73	0.88	0.74	0.96	0.72	0.87	0.39	0.42	0.42	0.42
11	Salt marsh	419	0.98	0.95	0.47	0.31	0.99	0.85	0.93	0.95	0.94	0.92	0.93	0.87
12	Mud flats	503	0.80	0.96	0.33	0.45	0.79	0.91	0.75	0.89	0.27	0.44	0.29	0.46
13	Water	927	0.98	1.00	0.93	0.97	0.98	0.99	0.98	0.98	0.94	0.92	0.96	0.98
	Average Acc. and Average Rel.		**0.80**	**0.78**	**0.45**	**0.45**	**0.77**	**0.76**	**0.73**	**0.72**	**0.57**	**0.55**	**0.56**	**0.55**
	Kappa coefficient		**0.81**		**0.46**		**0.78**		**0.77**		**0.59**		**0.59**	

Table 3. The values of parameter *Z* obtained by McNemar test.

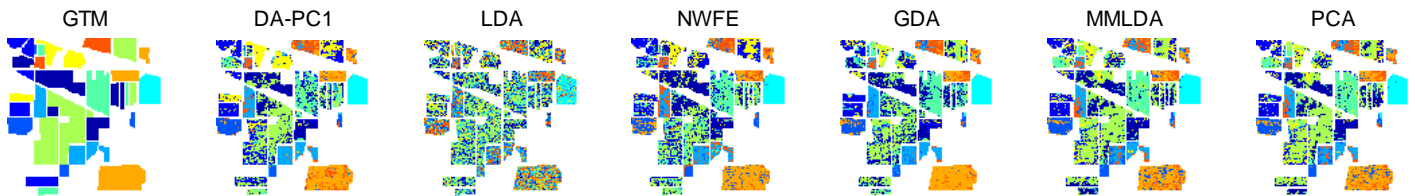| Indian/ SVM classifier/15 training samples/ 6 features							KSC/ SVM classifier/15 training samples/ 8 features					
	DA-PC1	LDA	NWFE	GDA	MMLDA	PCA		DA-PC1	LDA	NWFE	GDA	MMLDA	PCA
DA-PC1	0	48.92	15.65	6.15	9.63	9.27	DA-PC1	0	35.74	5.93	7.75	27.62	28.83
LDA	-48.92	0	-38.58	-44.26	-42.58	-42.77	LDA	-35.74	0	-32.18	-30.94	-13.80	-12.92
NWFE	-15.65	38.58	0	-8.40	-5.95	-6.33	NWFE	-5.93	32.18	0	2.00	22.44	23.59
GDA	-6.15	44.26	8.40	0	3.00	2.66	GDA	-7.75	30.94	-2.00	0	21.73	22.80
MMLDA	-9.63	42.58	5.95	-3.00	0	-1.23	MMLDA	-27.62	13.80	-22.44	-21.73	0	1.53
PCA	-9.27	42.77	6.33	-2.66	1.23	0	PCA	-28.83	12.92	-23.59	-22.80	-1.53	0

Figure 3. GTM and the classification maps of Indian dataset obtained by SVM classifier, 15 training samples and 6 extracted features.
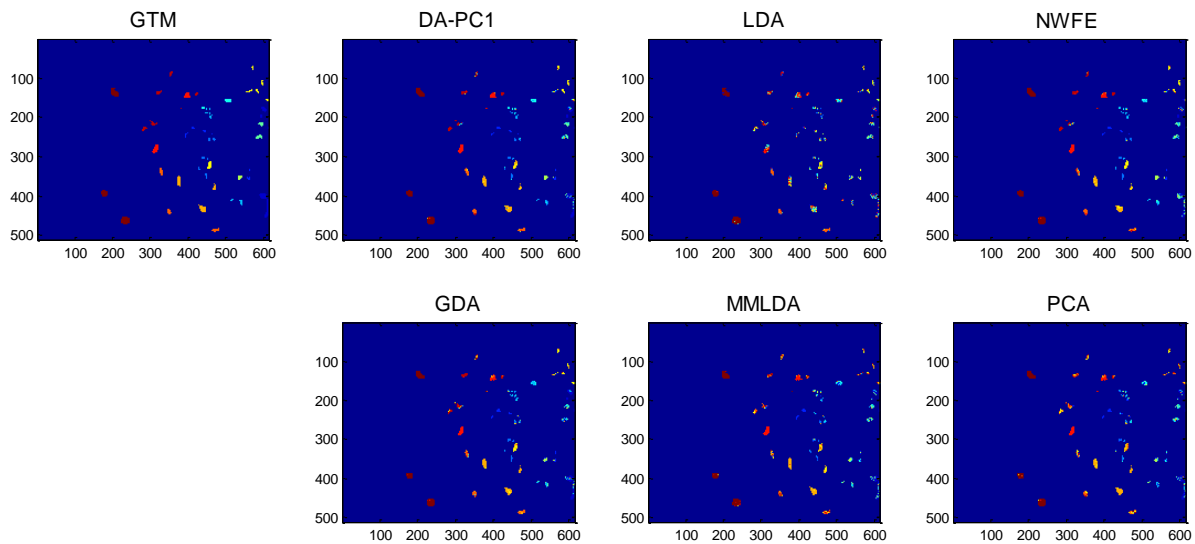

Figure 4. GTM and the classification maps of KSC dataset obtained by SVM classifier, 15 training samples and 8 extracted features.

Table 4. The highest classification accuracies achieved by 15 training samples.

| Dataset	Classifier	DA-PC1	LDA	NWFE	GDA	MMLDA	PCA
Indian	SVM	**0.72 (6)**	0.31 (8)	0.62 (6)	0.66 (5)	0.68 (4)	0.67 (4)
	ML	**0.72 (5)**	0.27 (6)	0.65 (6)	0.71 (4)	0.69 (4)	0.69 (4)
	NN	**0.72 (9)**	0.33 (8)	0.65 (11)	0.49 (6)	0.69 (4)	0.64 (4)
Pavia University	SVM	**0.79 (6)**	0.53 (6)	0.74 (12)	0.74 (6)	**0.79 (11)**	0.77 (11)
	ML	0.80 (6)	0.49 (5)	0.79 (3)	**0.81 (4)**	**0.81 (4)**	**0.81 (4)**
	NN	0.78 (6)	0.57 (5)	0.76 (10)	0.73 (8)	**0.80 (11)**	0.77 (5)
KSC	SVM	**0.80 (8)**	0.47 (10)	**0.80 (12)**	0.75 (10)	0.59 (15)	0.63 (15)
	ML	0.75 (6)	0.36 (6)	0.72 (8)	**0.77 (7)**	0.54 (3)	0.54 (3)
	NN	**0.79 (6)**	0.52 (12)	0.78 (11)	0.54 (7)	0.58 (15)	0.57 (14)
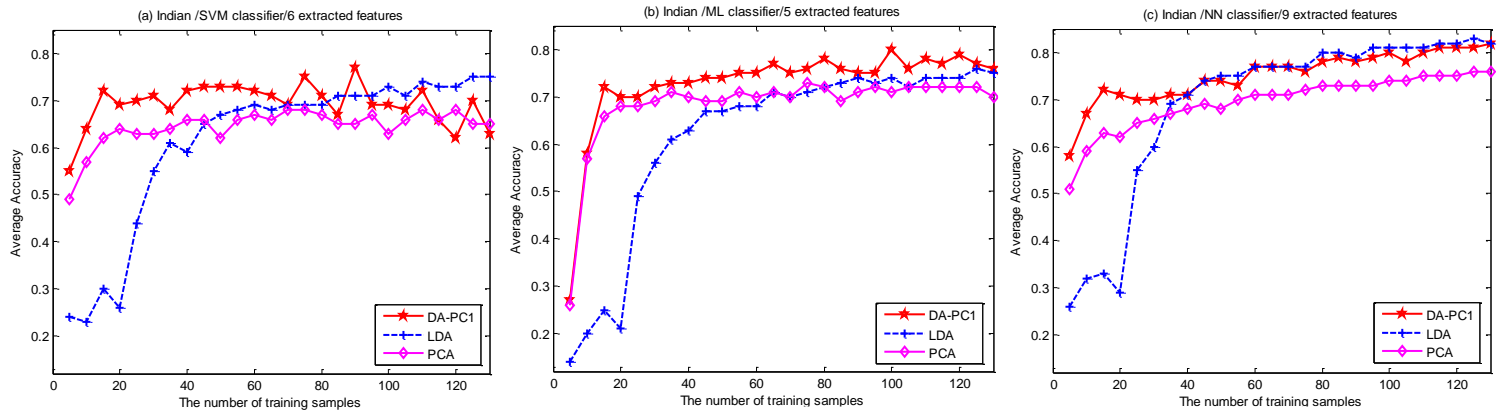
Figure 5. The comparison of performance of DA-PC1 with PCA and LDA in a fixed number of extracted features with varying the number of training samples for Indian dataset with a) SVM classifier and 6 extracted features, b) ML classifier and 5 extracted features, c) NN classifier and 9 extracted features.

References

[1] Richards, J. A. & Jia, X. (2006). Remote sensing digital image analysis, 4th Ed. New York: Springer Verlagl.

[2] Antonio, P., Jon, A. B., Joseph, W. B., Jason, B., Lorenzo, B., Gustavo, C., Jocelyn, C., Mathieu, F., Paolo, G., Anthony, G., Mattia, M., James, C. T. & Giovanna, T. (2009). Recent Advances in Techniques for Hyperspectral Image Processing. Remote Sensing of Environment, vol. 113, no.1, pp. S110-S122.

[3] Benediktsson, J. A., Pesaresi, M. & Amason, K. (2003). Classification and Feature Extraction for Remote Sensing Images From Urban Areas Based on Morphological Transformations. IEEE Transactions on Geoscience and Remote Sensing, vol. 41, no. 9, pp. 1940–1949.

[4] Yanfei, Z. & Liangpei, Z. (2012). An Adaptive Artificial Immune Network for Supervised Classification Of Multi-/Hyperspectral Remote Sensing Imagery. IEEE Transactions on Geoscience and Remote Sensing, vol. 50, no. 3, pp. 894–909.

[5] Moser, G. & Serpico, S. B. (2013). Combining Support Vector Machines and Markov Random Fields in an Integrated Framework for Contextual Image Classification. IEEE Transactions on Geoscience and Remote Sensing, vol. 51, no. 5, pp. 2734–2752.

[6] Demir, B. & Ertürk, S. (2009). Clustering-Based Extraction of Border Training Patterns for Accurate SVM Classification of Hyperspectral Images. IEEE Geoscience and Remote Sensing Letters, vol. 6, no. 4, pp. 840 - 844.

[7] Hughes, G. F. (1968). On The Mean Accuracy of Statistical Pattern Recognition. IEEE Transactions on Information Theory, vol. IT-14, no. 1, pp. 55–63.

[8] Korycinski, D. , Crawford, M. , Barnes, J. W. & Ghosh, J. (2003). Adaptive Feature Selection for Hyperspectral Data Analysis Using A Binary Hierarchical Classifier and Tabu Search. IEEE Symposium on Geoscience and Remote Sensing, Toulouse, France, 2003.

[9] Li, S., Wu, H., Wan, D. & Zhu, J. (2011). An Effective Feature Selection Method for Hyperspectral Image Classification Based on Genetic Algorithm and Support Vector Machine. Knowledge-Based Systems, vol. 24, no. 1, pp. 40-48.

[10] Yin, J., Wang, Y. & Hu, J. (2012). A New Dimensionality Reduction Algorithm for Hyperspectral Image Data Using Evolutionary Strategy. IEEE Transactions on Industrial Informatics, vol. 8, no. 4, pp. 935–943.

[11] Jia, X., Kuo, B. C. & Crawford, M. (2013). Feature Mining for Hyperspectral Image Classification. Proceedings of the IEEE, vol. 101, no. 3, pp. 676-697.

[12] Zhang, Q., Tian, Y., Yang, Y. & Pan, C. (2015). Automatic Spatial–Spectral Feature Selection for Hyperspectral Image via Discriminative Sparse Multimodal Learning. IEEE Transactions on Geoscience and Remote Sensing, vol. 53, no. 1, pp. 261-279.

[13] Li, S. & Wei, D. (2014). Extremely High Dimensional Feature Selection via Feature Generating Samplings. IEEE Transactions on Cybernetics, vol. 44, no. 6, pp. 737-747.

[14] Ladha, L. & Deepa, T. (2011). Feature Selection Methods and Algorithms. International Journal on Computer Science and Engineering, vol. 3, no. 5, pp. 1787–1797.

[15] Serpico, S. B. & Bruzzone, L. (2001). A New Search Algorithm for Feature Selection in Hyperspectral Remote Sensing Images. IEEE Transactions on Geoscience and Remote Sensing, vol. 39, no. 7, pp. 1360–1367.

[16] Peng, H., Long, F. & Ding, C. (2005). Feature Selection Based on Mutual Information: Criteria of Max Dependency, Max-Relevance, and Min-Redundancy. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, no. 8, pp. 1226-1238.

[17] Liu, X., Wang, L., Zhang, J., Yin, J. & Liu, H. (2014). Global and Local Structure Preservation for Feature Selection. IEEE Transactions on Neural Networks and Learning Systems, vol. 25, no. 6, pp. 1083–1095.

[18] Yuan, Y., Zhu, G. & Wang, Q. (2015). Hyperspectral Band Selection by Multitask Sparsity Pursuit. IEEE Transactions on Geoscience and Remote Sensing, vol. 53, no. 2, pp. 631–644.

[19] Imani, M. & Ghassemian, H. (2014). Band Clustering-Based Feature Extraction for Classification of Hyperspectral Images Using Limited Training Samples. IEEE Geoscience and Remote Sensing Letters, vol. 11, no. 8, 1325 – 1329.

[20] Jia, Y.-f., Li, Y.-j., Fu, P.-b. & Tian, Y. (2015). Nearest Feature Line and Point Embedding for Hyperspectral Image Classification. IEEE Geoscience and Remote Sensing Letters, vol. 12, no. 3, pp. 651-655.

[21] Kang, X., Li, S., Fang, L. & Benediktsson, J. A. (2015). Intrinsic Image Decomposition for Feature Extraction of Hyperspectral Images. IEEE Transactions on Geoscience and Remote Sensing, vol. 53, no. 4, pp. 2241-2253.

[22] Imani, M. & Ghassemian, H. (2014). Feature Extraction Using Attraction Points for Classification of Hyperspectral Images in a Small Sample Size Situation. IEEE Geoscience and Remote Sensing Letters, vol. 11, no. 11, pp. 1986 - 1990.

[23] Dhanjal, C., Gunn, S. R. & Shawe-Taylor, J. (2009). Efficient Sparse Kernel Feature Extraction Based on Partial Least Squares. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 31, no. 8, pp. 1347-1361.

[24] Kamandar, M. & Ghassemian, H. (2013). Linear Feature Extraction for Hyperspectral Images Based on Information Theoretic Learning. IEEE Geoscience and Remote Sensing Letters, vol. 10, no. 4, pp. 702 - 706.

[25] Xia, J., Chanussot, J., Du, P. & He, X. (2015). Spectral–Spatial Classification for Hyperspectral Data Using Rotation Forests with Local Feature Extraction and Markov Random Fields. IEEE Transactions on Geoscience and Remote Sensing, vol. 53, no. 5, pp. 2532–2546.

[26] Imani, M. & Ghassemian, H. (2015). Feature Space Discriminant Analysis for Hyperspectral Data Feature Reduction. ISPRS Journal of Photogrammetry and Remote Sensing, vol. 102, pp. 1 – 13.

[27] Fukunaga, K. (1990). Introduction to statistical pattern recognition. San Diego, CA. USA: Academic.

[28] Baudat, G. & Anouar, F. (2000). Generalized Discriminant Analysis Using a Kernel Approach. Neural Computation, vol. 12, pp. 2385–2404.

[29] Kuo, B. C. & Landgrebe, D. A. (2004). Nonparametric Weighted Feature Extraction for Classification. IEEE Transactions on Geoscience and Remote Sensing, vol. 42, no. 5, pp. 1096-1105.

[30] Xu, J., Yang, J., Gu, Z. & Zhang, N. (2014). Median–mean Line Based Discriminant Analysis. Neurocomputing, vol. 123, pp. 233–246.

[31] Cohen, J. (1960). A Coefficient of Agreement From Nominal Scales. Educational and Psychological Measurements, vol. 20, no. 1, pp. 37–46.

[32] Foody, G. M. (2004). Thematic Map Comparison: Evaluating the Statistical Significance of Differences in Classification Accuracy. Photogrammetric Engineering and Remote Sensing, vol. 70, no. 5, pp. 627–633.

کاهش ویژگی تصاویر ابرطیفی با استفاده از تحلیل ممیز و مولفه اصلی اول

مریم ایمانی و حسن قاسمیان*

دانشکده مهندسی برق و کامپیوتر، دانشگاه تربیت مدرس، تهران، ایران.

چکیده:

هنگامی‌که تعداد نمونه‌های آموزشی محدود است، کاهش ویژگی نقش مهمی در طبقه‌بندی تصاویر ابرطیفی ایفا می‌کند. در این مقاله یک روش استخراج ویژگی نظارت شده بر مبنای تحلیل ممیز پیشنهاد شده است که از مولفه‌ی اصلی اول برای وزن‌دهی ماتریس‌های پراکندگی استفاده می‌کند. روش پیشنهادی با مسئله‌ی نمونه‌ی آموزشی محدود مقابله می‌کند و محدودیت تحلیل ممیز خطی را در تعداد ویژگی‌های استخراجی ندارد. در روش پیشنهادی، ساختار غالب توزیع با مولفه‌ی اصلی اول حفظ می‌شود و جدایی‌پذیری کلاس‌ها با استفاده از تحلیل ممیز افزایش می‌یابد. نتایج آزمایش‌های انجام شده، نشان دهنده‌ی کارایی خوب روش پیشنهادی در مقایسه با سایر روش‌های استخراج ویژگی پرکاربرد است.

کلمات کلیدی: تحلیل ممیز، مولفه‌ی اصلی، کاهش ویژگی، ابرطیفی، طبقه‌بندی.