Dimensionality Reduction and Linear Discriminant Analysis for Hyperspectral Image Classification

Qian Du and Nicolas H. Younan

Department of Electrical and Computer Engineering
Mississippi State University, USA
{du,younan}@ece.msstate.edu

Abstract. In this paper, we investigate the application of Fisher's linear discriminant analysis (FLDA) to hyperspectral remote sensing image classification. The basic idea of FLDA is to design an optimal transform so that the classes can be well separated in the low-dimensional space. The practical difficulty of applying FLDA to hyperspectral images includes the unavailability of enough training samples and unknown information for all the classes present. So the original FLDA is modified to avoid the requirements of complete class knowledge, such as the number of actual classes present. We also investigate the performance of the class of principal component analysis (PCA) techniques prior to FLDA and find that the interference and noise adjusted PCA (INAPCA) can provide the improvement in the final classification.

Keywords: Fisher's Linear Discriminant Analysis, Dimensionality Reduction, Classification, Hyperspectral Imagery.

1 Introduction

Pattern recognition techniques have been widely used in remote sensing image analysis [1]. As a standard technique for dimensionality reduction in pattern recognition, Fisher's linear discriminant analysis (FLDA) projects the original high-dimensional data onto a low-dimensional space, where all the classes can be well separated [2]. Assume there are n training sample vectors given by $\{\mathbf{r}_i\}_{i=1}^{n}$ for p-classes $\{C_1, C_2, ..., C_p\}$, and there are n_j samples for the jth class, i.e., $\sum_{j=1}^{p} n_j = n$. Let $\boldsymbol{\mu}$ be the mean of the entire training samples, i.e., $\frac{1}{n}\sum_{i=1}^{n} \mathbf{r}_i = \boldsymbol{\mu}$, and $\boldsymbol{\mu}_j$ be the mean of the jth class, i.e., $\frac{1}{n_j}\sum_{\mathbf{r}_i \in C_j} \mathbf{r}_i = \boldsymbol{\mu}_j$. Then, the within-class scatter matrix $\mathbf{S}_W$ and between-class scatter matrix $\mathbf{S}_B$ are defined as

$$\mathbf{S}_W = \sum_{\mathbf{r}_i \in C_j} (\mathbf{r}_i - \boldsymbol{\mu}_j)(\mathbf{r}_i - \boldsymbol{\mu}_j)^T \tag{1}$$

I. Lovrek, R.J. Howlett, and L.C. Jain (Eds.): KES 2008, Part III, LNAI 5179, pp. 392 – 399, 2008.

$$\mathbf{S}_B = \sum_{j=1}^{p} n_j (\boldsymbol{\mu}_j - \boldsymbol{\mu})(\boldsymbol{\mu}_j - \boldsymbol{\mu})^T \tag{2}$$

respectively. The goal is to find a transform vector **w** such that the Raleigh quotient is maximized, which is defined as

$$q = \frac{\mathbf{w}^T \mathbf{S}_B \mathbf{w}}{\mathbf{w}^T \mathbf{S}_W \mathbf{w}} . \tag{3}$$

The **w** can be determined by solving a generalized eigen-problem specified by

$$\mathbf{S}_B \mathbf{w} = \lambda \mathbf{S}_W \mathbf{w} , \tag{4}$$

where λ is a generalized eigenvalue. Since the rank of $\mathbf{S}_B$ is p–1, there are p–1 eigenvectors associated with p–1 nonzero eigenvalues. Therefore, an $L \times (p-1)$ matrix **W** can be found to transform the original L-dimensional data into a (p–1)-dimensional space. In this low-dimensional space, it is expected that the p classes can be well separated.

The major problem when applying FLDA to remote sensing imagery is the difficulty in finding enough training samples for all the classes. In particular, for an L-band hyperspectral image, L linearly independent samples are required to make $\mathbf{S}_W$ full-rank. This problem may be resolved by a dimensionality reduction approach such as principal component analysis (PCA) before FLDA [3-5]. In this paper, we will also investigate other dimensionality reduction approaches, such as noise-adjusted principal component analysis (NAPCA) [6-7], interference and noise adjusted principal component analysis (INAPCA) [8], and their performance in conjunction with FLDA for hyperspectral image classification.

When applying FLDA, it is assumed that the number of classes, p, present in the image scene is known. For remote sensing imagery, this information may be difficult to know due to scene complexity. In many practical situations, we may know the number of foreground classes but not the number of background classes. We will investigate the performance of FLDA in this situation and propose an approach to overcome this difficulty.

2 PCA-Class Techniques for Dimensionality Reduction

2.1 PCA

Consider the observation model

$$\mathbf{z} = \mathbf{s} + \mathbf{n} , \tag{5}$$

where **z** is a pixel vector in a hyperspectral image with data dimensionality L (i.e, the number of spectral bands is L), **s** is a signal vector, and **n** represents the uncorrelated additive noise. Let $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_L]$ and $\boldsymbol{\Lambda} = diag\{\lambda_1, \lambda_2, \cdots, \lambda_L\}$ be the eigenvector and eigenvalue matrices of the data covariance matrix $\boldsymbol{\Sigma}$, where $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_L$ are

L eigenvectors of size $L \times 1$ and $\lambda_1, \lambda_2, \cdots, \lambda_L$ are the corresponding L eigenvalues, i.e.,

$$\mathbf{V}^T \mathbf{\Sigma} \mathbf{V} = \mathbf{\Lambda} . \tag{6}$$

Then, the PC images can be calculated from

$$\mathbf{z}_{PCA} = \mathbf{\Lambda}^{-1/2} \mathbf{V}^T (\mathbf{z} - \mathbf{m}), \tag{7}$$

where $\mathbf{m}$ is the mean vector. Assume $\lambda_1 \geq \lambda_2 \geq, \cdots, \geq \lambda_L$, the variances of the L PC images of the transformed data using $\mathbf{z}_{PCA}$ are $\lambda_1, \lambda_2, \cdots, \lambda_L$, respectively.

2.2 NAPCA

NAPCA ranks principal components (PCs) in terms of the signal-to-noise ratio (SNR). It can be performed with two steps. The first step conducts the noise-whitening to the original data, and the second step performs the ordinary PCA to the noise-whitened data. Since the noise variance is unity in the noise-whitened data, the resultant PCs are in the order of SNR. Let $\mathbf{\Sigma}_n$ be the noise covariance matrix and $\mathbf{F}$ be the noise-whitening matrix such that

$$\mathbf{F}^T \mathbf{\Sigma}_n \mathbf{F} = \mathbf{I} , \tag{8}$$

where $\mathbf{I}$ is the identity matrix. Transforming $\mathbf{\Sigma}$ by $\mathbf{F}$, i.e.,

$$\mathbf{F}^T \mathbf{\Sigma} \mathbf{F} = \mathbf{\Sigma}_{n\_adj} , \tag{9}$$

where $\mathbf{\Sigma}_{n\_adj}$ is the covariance matrix with the noise being whitened. Finding a matrix $\mathbf{G}$ such that

$$\mathbf{G}^T \mathbf{\Sigma}_{n\_adj} \mathbf{G} = \mathbf{I} . \tag{10}$$

Then, the operator for NAPCA can be constructed by

$$\mathbf{z}_{NAPCA} = \mathbf{G}^T \mathbf{F}^T (\mathbf{z} - \mathbf{m}). \tag{11}$$

The major difficulty in performing NAPCA is having an accurate noise covariance matrix $\mathbf{\Sigma}_n$, which is difficult in general. The following method is adopted in our research for its simplicity and effectiveness [9]. Let $\mathbf{\Sigma}$ be decomposed as

$$\mathbf{\Sigma} = \mathbf{DED} , \tag{12}$$

where $\mathbf{D} = diag\{\sigma_1, \sigma_2, \cdots, \sigma_L\}$ is a diagonal matrix with σ_l^2 being the diagonal elements of $\mathbf{\Sigma}$, which is the variance of the l-th original band, and $\mathbf{E}$ is the correlation coefficient matrix whose ij-th element represents the correlation coefficient between the i-th and j-th bands. Similarly, in analogy with the decomposition of $\mathbf{\Sigma}$, its inverse $\mathbf{\Sigma}^{-1}$ can be also decomposed as

$$\boldsymbol{\Sigma}^{-1} = \mathbf{D}_{\boldsymbol{\Sigma}^{-1}} \mathbf{E}_{\boldsymbol{\Sigma}^{-1}} \mathbf{D}_{\boldsymbol{\Sigma}^{-1}} \ , \tag{13}$$

where $\mathbf{D}_{\boldsymbol{\Sigma}^{-1}} = diag\{\zeta_1, \zeta_2, \cdots, \zeta_L\}$ is a diagonal matrix with ζ_l^2 being the diagonal elements of $\boldsymbol{\Sigma}^{-1}$ and $\mathbf{E}_{\boldsymbol{\Sigma}^{-1}}$ is a matrix similar to $\mathbf{E}$ with the diagonal elements being one and all the off-diagonal elements being within $(-1, 1)$. It turns out that ζ_l^2 is the reciprocal of a good noise variance estimate of the l-th band [9]. Therefore, the noise covariance matrix $\boldsymbol{\Sigma}_n$ can be estimated by $\boldsymbol{\Sigma}_n = diag\{\zeta_1^{-2}, \zeta_2^{-2}, \cdots, \zeta_L^{-2}\}$, which is a diagonal matrix.

2.3 INAPCA

INAPCA ranks PCs in terms of the signal-to-interference-plus-noise ratio (SINR), where the interference can be considered as any unwanted signals. Inspired by the two-step procedure of NAPCA, here we introduce a similar two-step procedure for INAPCA, which can be easily implemented without requiring prior interference information. Let $\boldsymbol{\Sigma}_{i+n}$ be the interference-noise covariance matrix. Then, a matrix $\mathbf{B}$ is determined via the eigen-decomposition of $\boldsymbol{\Sigma}_{i+n}$ such that

$$\mathbf{B}^T \boldsymbol{\Sigma}_{i+n} \mathbf{B} = \mathbf{I} \, , \tag{14}$$

which is used to transform the data as

$$\mathbf{B}^T \boldsymbol{\Sigma} \mathbf{B} = \boldsymbol{\Sigma}_{i+n\_adj} \, . \tag{15}$$

Here, $\boldsymbol{\Sigma}_{i+n\_adj}$ is the covariance matrix with interference and noise being whitened. If the resultant data is transformed by an ordinary PCA, then, the net effect is to order the PCs in terms of SINR. So, a matrix $\mathbf{C}$ can be found via the eigen-decomposition of $\boldsymbol{\Sigma}_{i+n\_adj}$ such that

$$\mathbf{C}^T \boldsymbol{\Sigma}_{i+n\_adj} \mathbf{C} = \mathbf{I} \, . \tag{16}$$

Then, the PC images using INAPCA can be computed by

$$\mathbf{z}_{INAPCA} = \mathbf{C}^T \mathbf{B}^T (\mathbf{z} - \mathbf{m}) . \tag{17}$$

$\boldsymbol{\Sigma}_{i+n}$ can be determined by estimating individual interference signatures using some unsupervised techniques. But $\boldsymbol{\Sigma}_{i+n}$ can be more easily computed via orthogonal subspace projection without the need of interference and noise estimation, which is explained as follows. Let the desired signature matrix be $\mathbf{S} = \lfloor \mathbf{s}_1 \ \mathbf{s}_2 \cdots \mathbf{s}_p \rfloor$ with p desired object signatures. In order to get a data set with interference and noise only, all these p signatures can be annihilated by projecting the original data onto the subspace that is orthogonal to these desired signals using projection operator $\mathbf{P}^{\perp}$, where

$\mathbf{P}^{\perp} = \mathbf{I} - \mathbf{S}\left(\mathbf{S}^T\mathbf{S}\right)^{-1}\mathbf{S}^T$. Let $\hat{\mathbf{z}} = \mathbf{P}^{\perp}\mathbf{z}$. Note that, only interference and noise are present in $\hat{\mathbf{z}}$. Then, $\mathbf{\Sigma}_{i+n}$ can be easily estimated by using the sample covariance matrix of $\hat{\mathbf{z}}$.

3 Variant of Fisher's Linear Discriminant Analysis

The total scatter matrix $\mathbf{S}_T$ is defined as

$$\mathbf{S}_T = \sum_{i=1}^{n} (\mathbf{r}_i - \boldsymbol{\mu})(\mathbf{r}_i - \boldsymbol{\mu})^T \tag{18}$$

and it can be related to $\mathbf{S}_W$ and $\mathbf{S}_B$ by [2]

$$\mathbf{S}_T = \mathbf{S}_W + \mathbf{S}_B . \tag{19}$$

So the maximization of Eq. (3) is equivalent to maximizing

$$q' = \frac{\mathbf{w}^T \mathbf{S}_B \mathbf{w}}{\mathbf{w}^T \mathbf{S}_T \mathbf{w}} . \tag{20}$$

Following the same idea of FLDA, the solution will be the eigenvectors of the generalized eigenproblem: $\mathbf{S}_B \mathbf{w} = \lambda \mathbf{S}_T \mathbf{w}$.

$\mathbf{S}_T$ in Eq. (20) can be replaced by the data covariance matrix $\mathbf{\Sigma}$, i.e.,

$$\hat{\mathbf{S}}_T = \mathbf{\Sigma} = \sum_{i=1}^{N} (\mathbf{r}_i - \mathbf{m})(\mathbf{r}_i - \mathbf{m})^T \quad , \tag{21}$$

where $\mathbf{m}$ is the sample mean of the entire data set with N pixels, i.e., $\frac{1}{N}\sum_{i=1}^{N} \mathbf{r}_i = \mathbf{m}$.

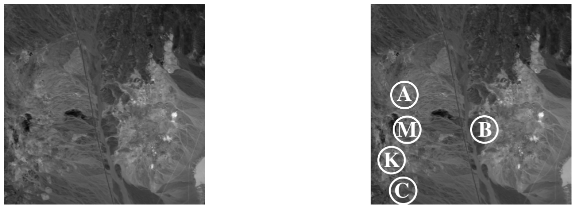
Then, the solution is the eigenvectors of the generalized eigenproblem: $\mathbf{S}_B \mathbf{w} = \lambda\, \mathbf{\Sigma} \mathbf{w}$ or $\mathbf{\Sigma}^{-1}\mathbf{S}_B$.

Regardless of the actual classes present in the data, replacing $\mathbf{S}_T$ with $\mathbf{\Sigma}$ represents an extreme case, which means all the pixels are separated into the classes they belong to and selected as samples. Here, the implicit assumption is that pixels are put into all the existing classes including unknown background classes (i.e., even when the actual number of classes, p_T, is unknown and greater than the p used in FLDA). In this paper, it will be shown that the term $\mathbf{\Sigma}^{-1}$ in this FLDA variant is very effective in background suppression so as to support the desired foreground classification.

4 Experiment

The AVIRIS Cuprite subimage scene of size 350×350, shown in Fig. 1, was collected in Nevada in 1997. The spatial resolution is 20m. Originally, it has 224 bands with

0.4-2.5 μm spectral range. After water absorption and low SNR bands were removed, 189 bands were used. This image scene is well understood mineralogically. At least five minerals were present: alunite (A), buddingtonite (B), calcite (C), kaolinite (K), and muscovite (M). Their approximate spatial locations of these minerals are marked in Fig. 1(b), but no pixel level ground truth is available. Due to the scene complexity, the actual number of classes, $p_{T,}$ is much greater than five. The number of training samples for the five classes was 63, 59, 69, 72, 63, respectively.



1(a) 1(b)

Fig. 1. AVIRIS Cuprite image scene (a) a spectral band image; (b) spatial locations of five pure pixels corresponding to minerals: alunite (A), buddingtonite (B), calcite (C), kaolinite (K), and muscovite (M)

Fig. 2(a) is the result using Target-Constrained Interference-Minimized Filter (TCIMF) on the original image [10], where the five classes were well separated. Fig. 2(b) is the result when TCIMF was applied to the FLDA-transformed images, where we can see there were a significant number of misclassified pixels, particularly when classified B, C, K, and M. Fig. 2(c), 2(d), and 2(e) are the results using PCA, NAPCA, and INAPCA before FLDA, where no obvious improvement can be perceived. This demonstrates that FLDA provides very poor results if the information on the number of classes is inaccurate. Under this circumstance, another dimensionality reduction process before FLDA cannot provide much help.

Fig. 2(f) is the result using the variant of FLDA, denoted as modified FLDA (MFLDA), where the classification is much better than that in Fig. 2(b) using the original FLDA. Fig. 2(g), 2(h), 2(i) are the results when using the three PCA-class techniques for dimensionality reduction before MFLDA and TCIMF. It seems that using INAPCA for dimensionality reduction, as illustrated in Fig. 2(i), may provide comparable results to that without dimensionality reduction, as shown in Fig. 2(f), while using PCA and NAPCA may not bring improvement.

To perform a quantitative comparison, Table 1 lists the spatial correlation coefficients between the results obtained from the use of LDA-transformed data and that from the TCIMF on the original data, where a value closer to one is associated with better classification. It is obvious that MFLDA outperforms the corresponding FLDA counterparts. This means $\mathbf{\Sigma}$ is a better term than $\mathbf{S}_W$ when the actual number of classes and their information are difficult or even impossible to obtain. It also confirms that using PCA and NAPCA for dimensionality reduction does not necessarily improve classification, but using INAPCA can improve classification particularly when MFLDA is applied.

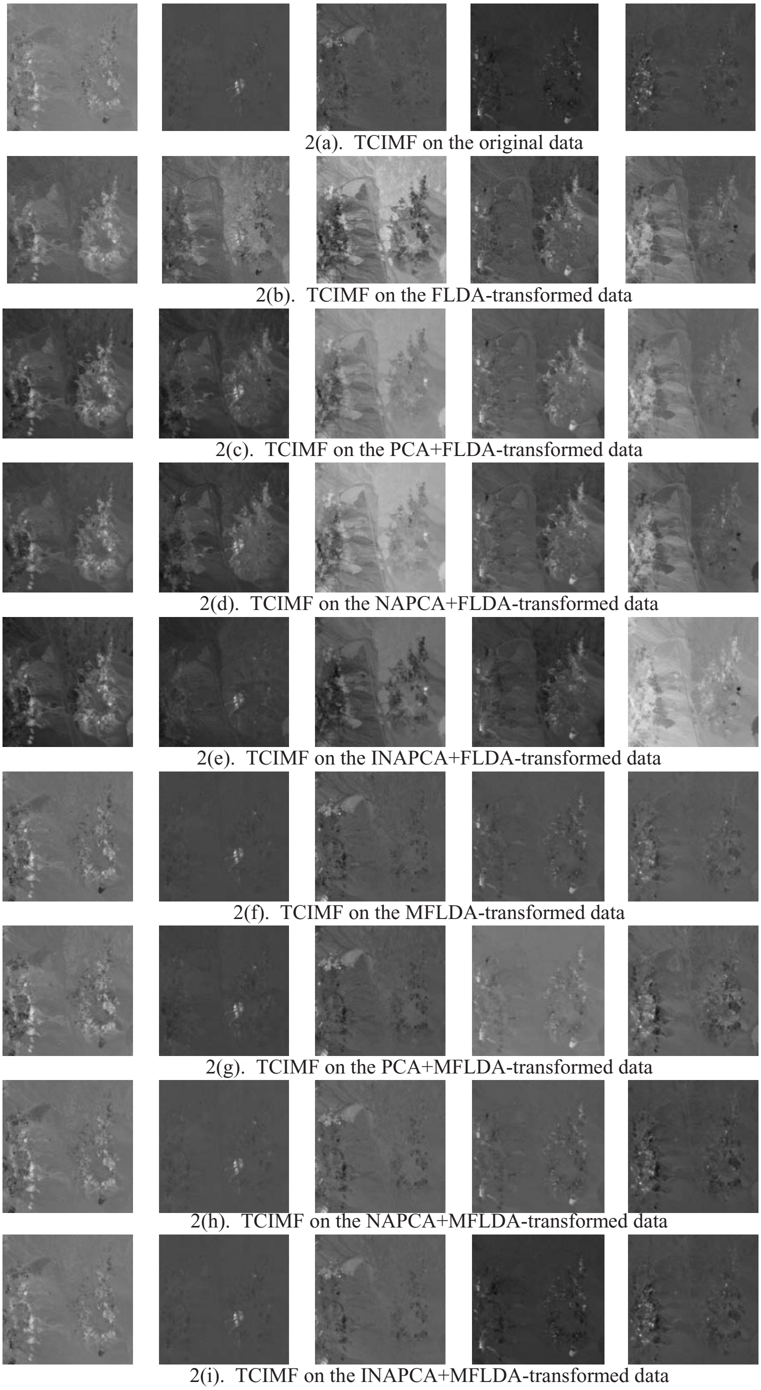


Fig. 2. Classification Results (left to right A, B, C, K, and M)

Table 1. Classification results compared with the TCIMF result using original data (correlation coefficient) in AVIRIS experiment

	A	B	C	K	M	Average
FLDA+TCIMF	0.49	0.16	0.08	0.49	0.13	0.27
PCA+FLDA+TCIMF	0.54	0.33	0.11	0.56	0.16	0.34
NAPCA+FLDA+TCIMF	0.54	0.30	0.11	0.55	0.14	0.33
INAPCA+FLDA+TCIMF	0.52	0.52	0.07	0.61	0.16	0.38
MFLDA+TCIMF	0.75	0.93	0.76	0.87	0.72	0.80
PCA+MFLDA+TCIMF	0.66	0.81	0.74	0.76	0.50	0.70
NAPCA+MFLDA+TCIMF	0.74	0.90	0.73	0.87	0.71	0.79
INAPCA+MFLDA+TCIMF	0.76	0.95	0.78	0.92	0.84	0.85

5 Conclusion

In this paper, we discuss how to effectively implement Fisher's linear discriminant analysis for remotely sensed hyperspectral imagery in the situation when the total number of classes is unknown. We find the variant, which uses the data covariance matrix Σ to replace the within-class scatter matrix $\mathbf{S}_{W,}$ can significantly improve the classification performance. In this case, using INAPCA as another dimensionality reduction approach prior to FLDA can further improve the classification performance, while the typical PCA or NAPCA cannot.

References

1. Landgrebe, D.A.: Signal Theory Methods in Multispectral Remote Sensing. John Wiley & Sons, Chichester (2003)
2. Duda, R.O., Hart, P.E.: Pattern Classification and Scene Analysis. John Wiley & Sons, New York (1973)
3. Belhumeur, P.N., Hespanha, J.P., Kriegman, D.J.: Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection. IEEE Transactions on Pattern Analysis and Machine Intelligence 19, 711–720 (1997)
4. Etemad, K., Chellappa, R.: Discriminant Analysis for Recognition of Human Face Images. Journal of Optical Society of America A 14, 1724–1733 (1997)
5. Casasent, D., Chen, X.-W.: Feature Reduction and Morphological Processing for Hyperspectral Image Data. Applied Optics 43, 227–236 (2004)
6. Green, A.A., Berman, M., Switzer, P., Craig, M.D.: A Transformation for Ordering Multispectral Data in Terms of Image Quality with Implications for Noise Removal. IEEE Transactions on Geoscience and Remote Sensing 26, 65–74 (1988)
7. Lee, J.B., Woodyatt, A.S., Berman, M.: Enhancement of High Spectral Resolution Remote Sensing Data by a Noise-Adjusted Principal Components Transform. IEEE Transactions on Geoscience and Remote Sensing 28, 295–304 (1990)
8. Chang, C.-I., Du, Q.: Interference and Noise-Adjusted Principal Components Analysis. IEEE Transactions on Geoscience and Remote Sensing 37, 2387–2396 (1999)
9. Roger, R.E., Arnold, J.F.: Reliably Estimating the Noise in AVIRIS Hyperspectral Imagers. International Journal of Remote Sensing 17, 1951–1962 (1996)
10. Ren, H., Chang, C.-I.: Target-Constrained Interference-Minimized Approach to Subpixel Target Detection for Hyperspectral Images. Optical Engineering 39, 3138–3145 (2000)