Submission Cover Sheet for 1999 IEEE SoutheastCon

Submission Type (Please Check One):

☐ *Full Length Paper - (Due 11/1/98)* ☑ *Concise Paper - (Due 12/1/98)*

Title of manuscript:
LINEAR DISCRIMINANT ANALYSIS FOR SIGNAL PROCESSING PROBLEMS

Authors and affiliations (List authors in the order to appear on the manuscript. List affiliations (companies, universities, etc.) and addresses exactly as they should appear. If there are more than 4 authors, please attach a second sheet.

1. Name **Suresh Balakrishnama**
Affiliation **Mississippi State University**
Address **PO Box 9571**
Mississippi State, MS 39762

2. Name **Aravind Ganapathiraju**
Affiliation **Mississippi State University**
Address **PO Box 9571**
Mississippi State, MS 39762

3. Name **Dr. Joseph Picone**
Affiliation **Mississippi State University**
Address **PO Box 9571**
Mississippi State, MS 39

4. Name
Affiliation
Address

Corresponding Author:
Name: **Suresh Balakrishnama**
Address: **PO Box 9571**
Mississippi State, MS 39762

Telephone: **601-325-8335**
FAX: **601-325-3149**
E-mail: **balakris@isip.msstate.edu**

Key Words: **classification, digital signal processing, linear discriminant analysis**

For Technical Committee Use Only -- Do Not Write Below

Date Received: Manuscript # Assigned

Author Notification Date: Date Camera Ready Received:

Decision: Session:

Notes:

LINEAR DISCRIMINANT ANALYSIS FOR SIGNAL PROCESSING PROBLEMS

Suresh Balakrishnama, Aravind Ganapathiraju, Joseph Picone

Institute for Signal and Information Processing
Mississippi State University
Mississippi State, Mississippi 39762, USA
email: {balakris, ganapath, picone}@isip.msstate.edu
Ph (601) 325 -3149 Fax (601) 325-3149

***Abstract* - Linear Discriminant Analysis (LDA) and Principal Components Analysis (PCA) are two common techniques used for classification and dimensionality reduction. These techniques typically use a linear transformation which can either be implemented in a class-dependent or class-independent fashion. PCA is a feature classification technique in which the data in the input space is transformed to a feature space where the features are decorrelated. On the other hand, the optimization criterion for LDA attempts to maximize class separability. In this paper we quantify the efficacy of these two algorithms along with two other classification techniques, Support Vector Machines (SVM) and Independent Components Analysis (ICA). The problem of classifying forestry images based on their scenic beauty is considered. On a standard evaluation task consisting of 478 training images and 159 test images, class-dependent LDA produced a 35.22% misclassification rate, which is significantly better than the 43.3% rate obtained using PCA and is on par with the performance of ICA and SVM.**

INTRODUCTION

The Euclidean distance metric is the simplest and most commonly used distance metric for classification of data. It is assumed that we can define a representative member for every class involved in the classification problem. The test sample is then assigned the class to whose representative member it is closest. However, classes may have significant overlap in the input space making the Euclidean distance metric ineffective. To counter this problem, data is transformed to a feature space where the classes are better separated or the features are independent and the distance metric is more effective. Figure 1 shows a typical sequence of steps employed in the implementation of a classifier. The signal modeling techniques discussed in this paper help define the new feature space.

Linear Discriminant Analysis (LDA), like Principal Components Analysis (PCA), is a tool for multigroup data classification and dimensionality reduction. LDA maximizes the ratio of between-class variance to the within-class variance in any particular data set thereby guaranteeing maximal separability [1]. The primary difference between LDA and PCA is that PCA performs feature classification while LDA performs data classification. PCA changes both the shape and location of the data in its transformed space whereas LDA builds a definite decision region between the classes.

Independent Components Analysis (ICA) [5] and Support Vector Machines (SVM) [6] are two nonlinear classification techniques getting significant attention in recent years. There is, however, a clear distinction between SVM and other techniques discussed in this paper. SVM is a classification technique whereas the others are data modeling techniques that make classification more effective. ICA and SVM are inherently iterative in nature and, computationally, are very expensive compared to LDA or PCA. However, they have been applied to several classification tasks with extremely encouraging results [6], performing better than other nonlinear classifiers like neural networks and radial basis functions. In this paper we describe LDA in detail and present results of the different classification schemes on the forestry image data.

LINEAR DISCRIMINANT ANALYSIS

LDA is a transform-based method which attempts to minimize the ratio of within-class scatter to the between-class scatter. The mathematical formulation involved in the theory of LDA is explained in the following sections.

A within-class scatter matrix defines the scatter of samples around their respective class centers (means) and is

Figure 1. Overview of the classification problem showing basic components involved in a classifier

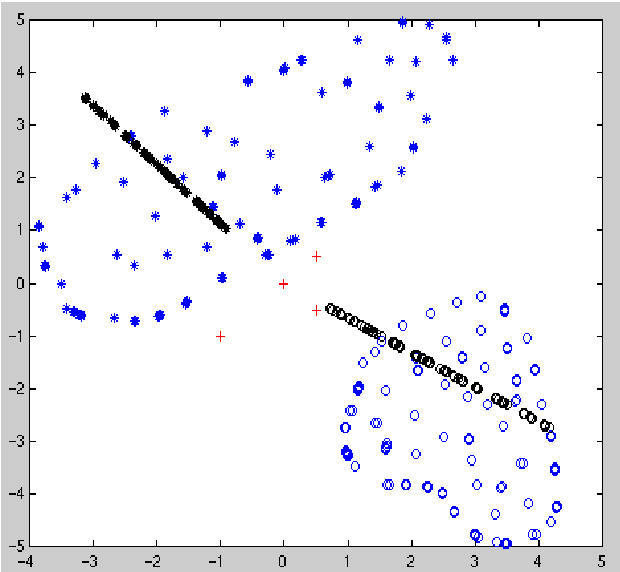

Figure 2. An example of class-dependent LDA with the data in the original space (circles), data in transformed space (diagonal lines), test points (crosses) and the decision region obtained.

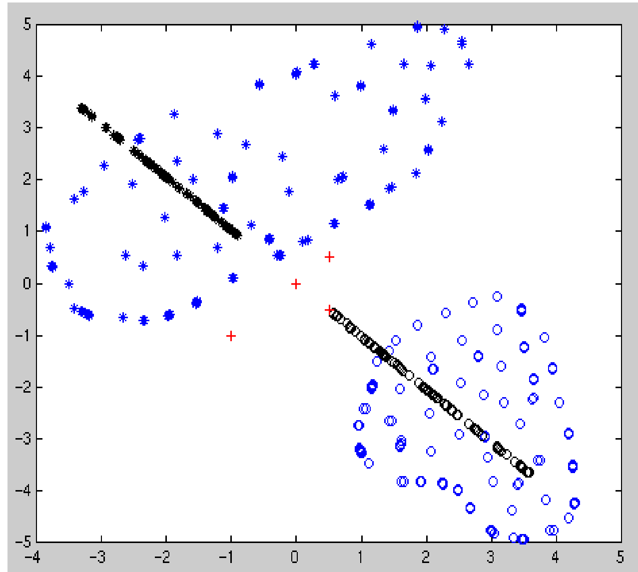

Figure 3. An example of class-independent LDA with the data in the original space (circles), data in transformed space (diagonal lines), test points (crosses) and the decision region obtained.

computed using

$$S_w = \sum_{i=1}^{L} P_i E\left\{(X - M_i)(X - M_i)^T\right\}, \qquad (1)$$

where M_i is the mean of the i'th class and P_i is the relative occurrence of members of class i in the training data. A between-class scatter matrix defines the scatter of the expected vectors around the global mean and is computed as

$$S_b = \sum_{i=1}^{L} P_i (M_i - M_o)(M_i - M_o)^T \qquad (2)$$

where, M_o is the global mean, M_i is the mean of the i'th class and P_i is the relative occurrence of members of class i in the training data.

The overall or mixture scatter matrix is obtained by the covariance matrix of all samples and is computed using Equation 3.

$$S_m = E\left\{(X - M_o)(X - M_o)^T\right\} = S_w + S_b \qquad (3)$$

The optimizing criterion to obtain the LDA transform is a combination of within-class scatter, between-class scatter and the mixture-scatter. The criterion is set to maximize the separation of the classes in the transformed space. The criteria commonly used are:

$$criterion = \ln\left|inv(S_w) \times S_b\right| \qquad (4)$$

$$criterion = S_b - \mu(S_w - c) \qquad (5)$$

$$criterion = inv(S_w) \times S_b \qquad (6)$$

LDA uses the last of these to optimize the class separation in the transform space.

The transformation matrix is formed by the eigenvectors corresponding to the dominant eigenvalues of the optimizing criterion. An eigenvector of a transformation represents a 1-D invariant subspace of the vector space in which the transformation is applied. A set of these eigenvectors whose corresponding eigenvalues are non-zero are all linearly independent and invariant. Thus, any vector space can be represented in terms of linear combinations of the eigenvectors. A linear dependency between features is indicated by a zero eigenvalue. The eigenvector is computed using

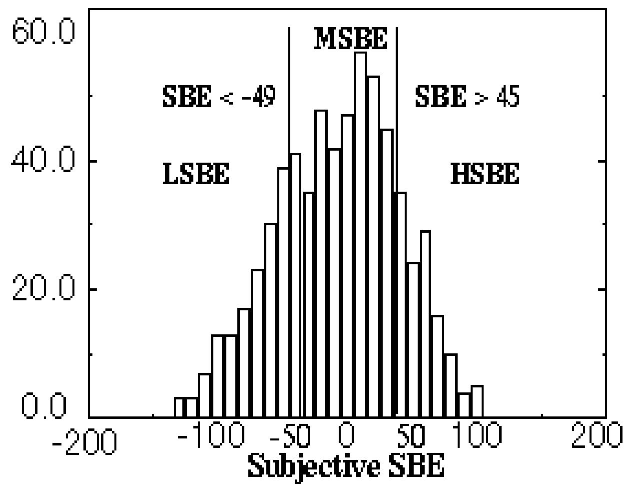

Figure 4. Distribution of USFS database according to the scenic beauty estimate ratings.

$$\Phi_1 = \frac{S_w^{-1}(M_i - M_j)}{\left\| S_w^{-1}(M_i - M_j) \right\|} \tag{7}$$

where, i and j correspond to different classes.

Once the transforms are obtained, the training data and test data are transformed to the new space. Based on the Euclidean distance of the transformed test sample from each of the class centres in the transformed space, a class is assigned to each test vector.

LDA for data classification can be implemented in two forms: *Class-dependent and Class-independent* transformations. The class-dependent approach involves maximizing the ratio of between-class covariance to within-class covariance for each class separately. This results in L transformations, each corresponding to one class. The class-independent approach involves maximizing the ratio of between-class scatter to the within-class scatter across all classes simultaneously. In the class-independent approach the optimizing criterion is used to define a single transformation. Figures 2 and 3 show examples of class-dependent LDA and class-independent LDA for a simple two-class problem.

APPLICATION TO SBE ESTIMATION

We have applied LDA to the task of classifying forestry images based on their scenic beauty [3]. The objective is to automatically classify a database of 637 images into either of three classes: high scenic beauty (HSBE), medium scenic beauty (MSBE) or low scenic beauty (LSBE). The reference values for the entire database are obtained by quantifying human judgements. Though the human subjects rated the images on a continuous scale, the problem was cast as a classification problem by dividing the range of the beauty estimates into three groups. Figure 4 shows how each image was assigned to a class based on the scenic beauty estimate (SBE) values [4].

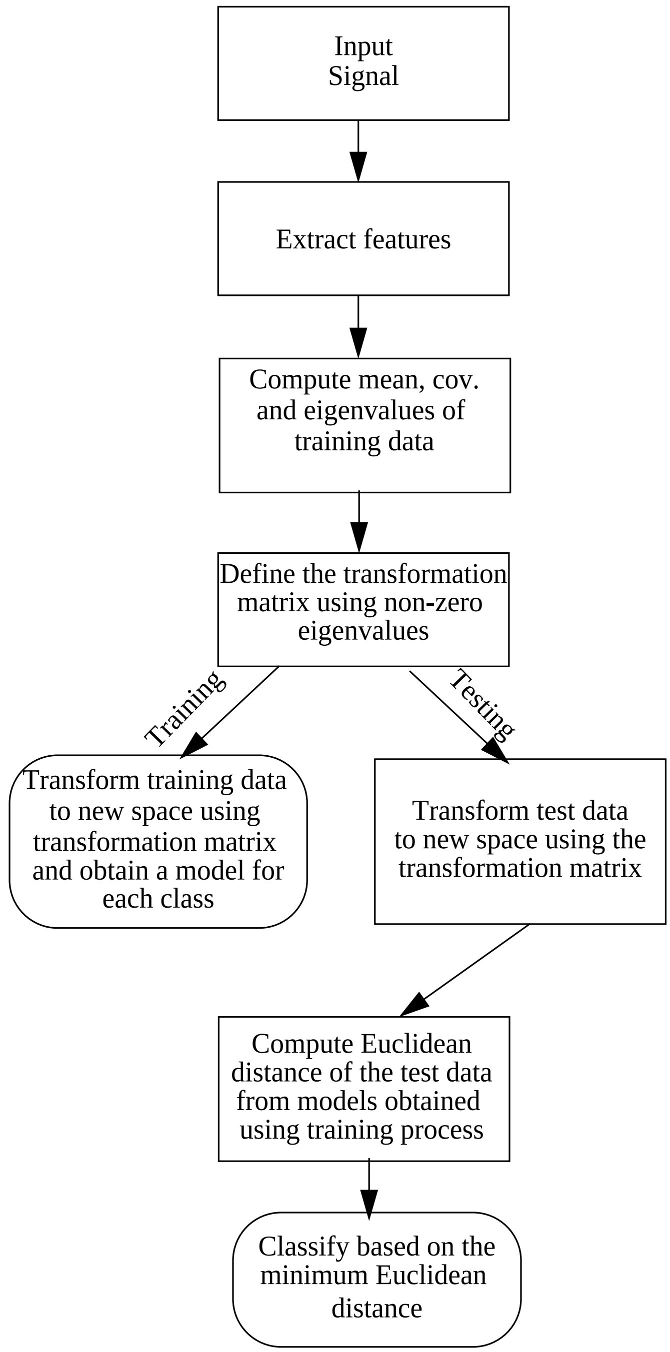

Figure 5. A block diagram showing the steps involved in classification using LDA.

Algorithm	Mis-classification Rate
LDA	35.22 %
PCA	43.30 %
ICA	33.44 %
SVM	33.00 %

Table 1. The performance of various classification algorithms on the forestry images.

Figure 5 outlines the process of training a system to perform classification of the forestry images. We use 45 features composed of red, green, and blue distributions, the number of long and short lines, entropy, compression ratio, and fractal dimension. The number of long and short lines act as a measure of the number of trees and bushes in the image. The distributions of red, blue and green are quantized to ten bins each.

EXPERIMENTS

Class-dependent LDA has been applied to classify the forestry images. The results obtained are compared with other classification algorithms like PCA, ICA and SVM. 478 training and 159 test images, constituting set 1 of the standard USFS database [4], were used.

The experimental setup consists of three phases: training, training-update and testing. The training process involves generating models for HSBE, MSBE and LSBE images of the training set. The mean and covariance of the features for each class are computed (these comprise the model definition). In the training-update phase between-class scatter and within-class scatter matrices of the training images are computed. The LDA transformation matrix is defined. The models for the three classes are transformed using this transformation matrix. In the testing phase, each test image is transformed using the transformation matrix for each of the classes. The distance of each test image from the model for each class is computed. Classes are assigned based on the minimum Euclidean distance. The efficiency of the classification is represented in terms of the misclassification rate. The misclassification rate is defined as the ratio of the number of misclassified images to the total number of test images.

Table 1 shows the relative performance of the four classification techniques discussed in this paper. It is clear from the results that data-classification techniques do a better job than feature-classification techniques like PCA. The highly nonlinear nature of this classification problem makes the nonlinear classifiers like SVM well suited for classification. However, the trade-off between computational costs and performance make LDA an attractive alternative.

CONCLUSIONS AND FUTURE WORK

In this work we have found that LDA performs on par with more advanced classification techniques. Though these other techniques yield slightly better results, LDA is attractive due to its computational simplicity. LDA was also shown to be superior to PCA for the task of classifying forestry images. As part of our future work, we plan to develop a Java-based demonstration which could be used to visualize LDA based transformations on user defined data sets and also help the user appreciate the difference between the various classification techniques.

REFERENCES

[1] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, Academic Press, San Diego, California, USA, 1990.

[2] S. Balakrishnama and A. Ganapathiraju, “Linear Discriminant Analysis - A brief Tutorial,” *http://www.isip.msstate.edu/publications/reports/*, Institute for Signal and Information Processing, Mississippi State University, MS State, MS, USA, March 2, 1998.

[3] N. Kalidindi and J. Picone, “Scenic Beauty Estimation of Forestry Images,” *USFS Project Final Report*, Institute for Signal and Information Processing, Mississippi State University, MS State, MS, USA, December 1997.

[4] N. Kalidindi, A. Le, J. Picone and V. A. Rudis, “USFS Scenic Beauty Estimation Database,” *USFS Southern Research Station*, Institute for Signal and Information Processing, Mississippi State University, MS State, MS, USA, December 1996.

[5] A. J. Bell and T. J. Sejnowski, “An Information-Maximization Approach to Blind Separation and Blind Deconvolution,” *Technical Report* No. INC-9501, Institute for Neural Computing, University of California at San Diego, San Diego, California, USA, February 1995.

[6] V. Vapnik, *The Nature of Statistical Learning Theory*, Springer-Verlag, New York, New York, USA, 1995.