

DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

*paper for*

**Scenic Beauty Estimation  
using  
Linear Discriminant Analysis**

*submitted to fulfill the semester project requirement for*

**EE 4773/6773: Digital Signal Processing**

**October 25, 1997**

*submitted to:*

**Dr. Joseph Picone**

**Department of Electrical and Computer Engineering  
413 Simrall, Hardy Rd.  
Mississippi State University  
Box 9571  
MS State, MS 39762**

*submitted by:*

**Suresh Babu Balakrishnama & Robert M. Brown Jr.  
Scenic Beauty Group  
Department of Electrical and Computer Engineering  
Mississippi State University  
Box 6176  
Mississippi State, Mississippi 39762  
Tel: 601-325-8335  
Fax: 601-325-3149  
email: balakris@isip.msstate.edu, rmb6@ra.msstate.edu**



**ABSTRACT**

The main focus of this project was to estimate the scenic beauty of forest images using Linear Discriminant Analysis, a valuable tool for multigroup classification and data reduction. This project originated from the United States Forest Service (USFS) to determine the scenic beauty of forests to preserve recreation and aesthetic resources in forest management. The results of this project will be useful to help determine a predefined pattern to cut trees so as to retain the scenic beauty even after cutting the forest for timber. The algorithm was initially developed and tested in Matlab, and the final algorithm was developed in C++. The software developed will be tested on 638 images available in the database to determine their scenic quality. Each image will be rated on a scale of -200 to 200 and the results obtained will be compared with the ratings from the Scenic Beauty Estimation human subjective ratings through the use of histograms.

## **Background**

The database to be used is based on two sets of images taken approximately four years apart. The first set consists of photographs taken during 1990-91. The second set consists of photographs taken during 1994-95. The images included in the database were taken from a study spanning four years in the Ouachita Forest in Arkansas. The photographs were taken under controlled conditions and digitized using extremely high quality scanning process.

The original images were delivered in a proprietary format developed by Kodak know as the PhotoCD format (PCD), a format for archiving high-quality photographs. Due to the proprietary nature of the PCD format, the images were converted to a Portable Pixel Map image (PPM). The PCD to PPM conversion was done using pcdtoppm version 0.6 on a Unix machine.

The PPM images were converted from a proprietary format developed by Kodak know as the PhotoCD format (PCD) with a resolution of 4x PCD to 1536 pixel wide by 1024 pixel high format of PPM. Each image requires about 5 Mbytes of disk space. The PPM P6 formatted image used represents each pixel as a sequence of three bytes. The first byte corresponds to the value of the color red encoded on a linear scale of 0 to 255. The second and third numbers represent the values for green and blue colors. Each file also includes information about the image such as the photographic tray and slide number, a human subjective judgement of 1 to 10, a Scenic Beauty Rating (SBE), the Normalized Scenic Beauty Rating (SBEZ), the block number, treatment, plot number, date, angle, and other miscellaneous informations.

## **Why LDA?**

There are many possible way to aid in classification of data, Principle Component Analysis (PCA) and Linear Discriminant Analysis (LDA) are two possible methods to use for data classification and reduction. We decided to implement an algorithm for LDA in hopes to provide better classification for images being tested. This method was also used to help better understand the distribution of the feature data. The success of LDA will help show how the data is distributed.

Linear Discriminant Analysis easily handles the case where the within class frequencies are unequal and their performances has been examined on randomly generated test data. This method maximizes the ratio of overall variance to the within class variance in any data class.

## **APPLYING LDA TO THE SCENIC BEAUTY ESTIMATION PROJECT**

Our LDA algorithm will be applied to a compilation of results computed from existing estimation algorithms in hopes to reduce the number of features to the most significant features. Our algorithm will be added to the pre-existing framework of the Scenic Beauty Estimation program. This will allow us to make use of the developed methods of data handling and evaluation processes.

The program will allow a user to specify the number of features to reduce the existing features to. The main program computes 42 different features that will be used in the LDA algorithm to determine correlation of the features and the image classification.

## **ALGORITHM TESTING**

The LDA technique will be initially coded in Matlab and tested with the training and testing data consisting of 42 features and results would be obtained by reducing features and the results (error performance) will be compared to the previous test results and human SBE ratings. These tests will provide a new space with reduced features where adequate discrimination information is available. This technique will also be applied on sets of synthetic data with 2 or 3 features. The synthetic data to be used will be data that fits the LDA model of specifically grouped data for training and data points inside and outside of the classes to be tested.

The LDA algorithm developed will be extensively tested in both its forms, Matlab and C++. The algorithm will be tested using synthetic data using the following scenarios. Data comprised of a two feature space (two dimensional data:  $x$  and  $y$ ) with three distinct classes, data comprised of a three feature space (three dimensional data:  $x$ ,  $y$ , and  $z$ ) with three distinct classes, data comprised of a  $n$  feature space ( $n$  dimensional data) with three distinct classes where the data will be computed using a Gaussian distribution to ensure that the algorithm will work properly with higher order data.

The final test set will be comprised of a two feature space with three overlapping classes of data. This data set will not have distinct classes of data except for those drawn for the purpose of defining classes. This data will not be able to be accurately classified with the use of LDA since it is not in distinct classes. This test is to confirm that our algorithm will not properly classify data that does not have distinct difference between classes.

## **PROJECT SUMMARY**

The project of Scenic Beauty Estimation using Linear Discriminant Analysis is to provide a more accurate estimation of beauty of images from the data base. This project is a continuation of work done as a part of work done by the Institute for Signal and Information Processing and projects done in EE4773/6773: Digital Signal Processing.

Linear Discriminant Analysis will be applied to the results of estimations developed from the pre-existing code. The algorithm was integrated into the existing Scenic Beauty Estimation code. This will allow further development of the Scenic Beauty Estimation project by providing another method data reduction for classification of images.







**VII. REFERECNES**

- [1] Nirmala Kalidindi, Audrey Le, Dr. Joe Picone, Dr. Victor A. Rudis, Scenic Beauty Estimation, Research Paper, Mississippi State University, 1996.
- [2] Nagendra Kumar, Ph.D. Thesis. John Hopkins University, 1996.
- [3] N. Kumar, A. Andreou, "On Generalizations of Linear Discriminant Analysis." Research Paper, John Hopkins University, 1996.
- [4] Richard O. Duda, Peter E. Hart, Pattern Classification and Scene Analysis. John Wiley & Sons, Inc., 1973.