Scenic Beauty Estimation Using Linear Discriminant Analysis

Robert M. Brown Jr., Suresh Babu Balakrisnama

Scenic Beauty Estimation Group
Department of Electrical and Computer Engineering
Mississippi State University
Box 9571
Mississippi State, MS 39762
434 Simrall, Hardy Rd.

rmb6@ra.msstate.edu, balakris@isip.msstate.edu

Abstract

The main focus of this project was to estimate the scenic beauty of forest images using Linear Discriminant Analysis, a tool for multigroup classification and data reduction. This project originated from the United States Forest Service (USFS) to determine the scenic beauty of forests and 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.

I. Introduction

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. [1]

The areas for study were partitioned into four blocks, with each block subdivided into five plots. The database includes images from each plot during each season of the year. The plots were photographed both in the 1990-91 period and the 1994-95 period. Each plot was photographed from at least four different angles. The total number of images in the database is 4 blocks x 5 plots x 4 seasons x 4 angles x 2 years = 640. An additional 40 images (20 images from the 1990-91 and 20 images from the 1994-95 period) were added as baseline images. A set of 20 images are also included as warm-up slides. These warm-up slides were used to calibrate human performance during subjective testing. [1]

Of the original 700 images received, they were partitioned into a set of 679 images used for development, 20 warm-up slides held out from the database proper, and one anomalous slide suffered from incomplete documentation. [1]

The original images were delivered in a proprietary format developed by Kodak and is 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. [1]

The PPM images were converted from a 4x PCD resolution to 1536 pixel wide x 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 to10, a Scenic Beauty Rating (SBE), the Normalized Scenic Beauty Rating (SBEZ), the block number, treatment, plot number, date, angle, and other miscellaneous information. [1] For the bulk of the project, we refer to the images in three general classes least scenic beauty estimate (LSBE), medium scenic beauty estimate (MSBE), and high scenic beauty estimate (HSBE).

Why Linear Discriminant Analysis?

There are many possible ways to aid in classification of data. Principle Component Analysis (PCA) and Linear Discriminant Analysis (LDA) are two possible methods used for data classification and data reduction. [6] 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 particular data class. [2][10] The use of Linear Discriminant Analysis will be applied to the existing results of the tests applied to the images in hopes to determine a more accurate overall rating for each image. We decided to implement an algorithm for LDA in hopes of providing better classification for images being tested. This method will also help to better understand the distribution of the feature data.

The following is a discussion of three examples for different types of data distributions and how the application of LDA and PCA can provide discrimination for the classes.

In Figure 1, the two class distribution shows the within-class distributions as Guassians with equal variance in a two dimensional sample space. The ellipses in the figure represent contours of equal probability density for the two Guassians. The direction of maximum variance for each of the Guassians is shown by the line labeled PCA. In this example, it is also in the direction of the maximum variance of the mixture of the two Guassians, and hence in the direction of the first principal component. Unfortunately, the projection on this line contains no class discrimination information. One optimum solution for this type of problem is to use methods of Linear Discriminant Analysis (LDA). The first linear discriminant is shown by the lines labeled LDA. The direction of LDA clearly demonstrates the discrimination maximallybetween the classes. [2]

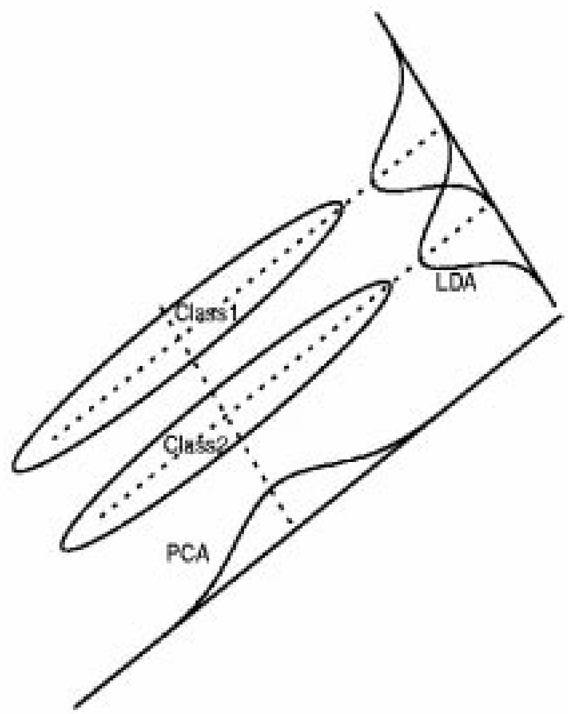

Figure 1: LDA succeeds and PCA fails in two Gaussian classes. [2]

For second example, shown Figure 2, the distribution of data is Guassian with unequal variance in a two dimensional sample space. The contours shown represent unequal probability density for the different Guassians. The direction of maximum variance for each of the Guassians is shown by the curves labeled PCA. In this case, PCA succeeds to show a projection that contains class discrimination information. A projection for LDA would also show class discrimination information. In this case, both PCA and LDA would be possible choices for class determination.

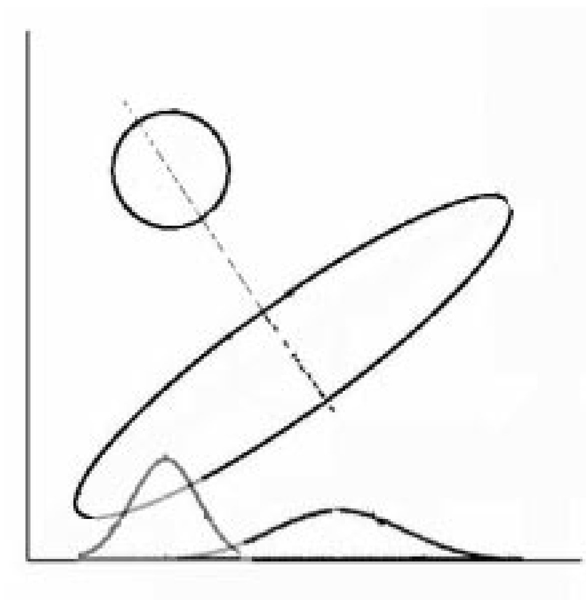

Figure 2: PCA succeeds in two Gaussian classes.

In final example, illustrated in Figure 3, the classes are shown with heteroscedastic within class distributions. In this example, the classes have similar means and drastically different variances. The projection for PCA is slightly better than that of LDA, but both projections show how LDA and PCA would probably fail in their ability to discriminate between the classes.[3]

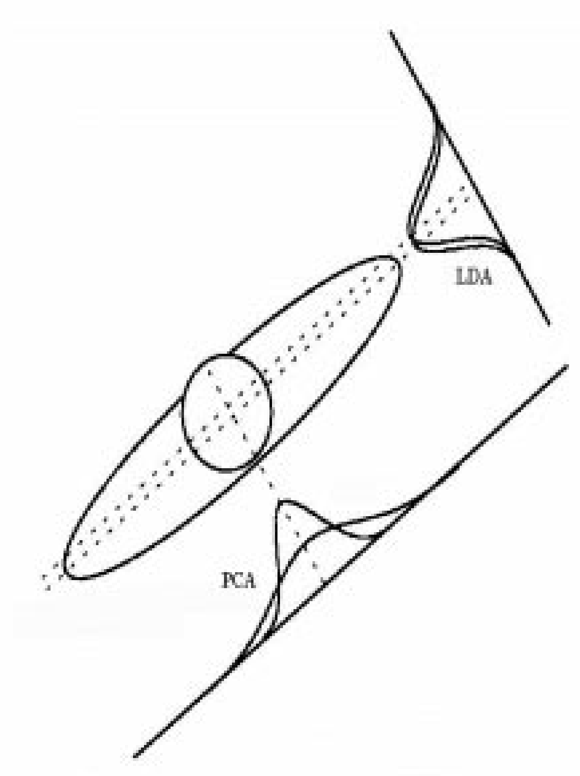

Figure 3: LDA and PCA fail in two Gaussian classes [2]

In this case, both projections for LDA and PCA provide minimal class discrimination. The project for PCA provides better class discrimination than that of LDA. In this case, neither LDA nor PCA provide adequate class discrimination to all accurate classification of data in a project such as ours.

II. THEORY

The problem of dimensionality reduction through linear projections is described as follows: Let x be an n dimensional feature vector. We seek a linear transformation $\Re^n \to \Re^p (p < n)$ of the form $y_p = \theta_p^T x$ where θ_p is a $n \times p$ matrix. Let θ be a non-singular $n \times n$ matrix used to define the linear transformation $y = \theta^T x$. Let us partition as

$$\theta = [\theta_p \theta_{n-p}] = [\vec{\theta}_1 \ldots \vec{\theta}_n] \quad (22)$$

where θ_p consists of the first p columns of θ and θ_{n-p} consists of the remaining $n-p$ columns and $\vec{\theta}_i$ is the i^{th} column of θ. Then, feature dimension reduction can be viewed as a two step procedure. First, a non-singular linear transformation is applied to x to obtain $y = \theta^T x$. Then, in the second step, only the first p rows of y are retained to give y_p. This notation may seem superfluous at the moment. [3]

The objective of LDA is to choose the linear transformation θ_p in such a way so as to retain the maximum amount of class discrimination information in the reduced feature space. Let there be a total of J classes, and let $g(i) \rightarrow \{1 \ldots J\}$ indicate the class that is associated with x_i. Let $\{x_i\}$ be the set of training examples available. Then $\sum_{g\langle i\rangle = j} 1 = N_j$ (here $g(i) = j$ is a notation used to define the set of all i for which $g(i) = j$) is the total number of training examples associated with class j, $\sum_{j=1}^{J} N_j = N$ is the total number of training examples. [2]

Let $\bar{X}$ be the sample mean, which is defined as

$$\bar{X} = \frac{1}{N} \sum_{i=1}^{N} x_i \quad (23)$$

The total normalized sum of squares and products (SSQP) T is defined as

$$\bar{T} = \frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{X})(x_i - \bar{X})^T \quad (24)$$

The class means $\overline{X_j}$, the normalized class SSQP $\overline{W_j}$, and the overall pooled and normalized SSQP $\overline{W}$ are defined as

$$\bar{X}_j = \frac{1}{N_j} \sum_{g(i) = j} x_i, j = 1 \ldots J \quad (25)$$

$$\overline{W_j} = \frac{1}{N_j} \sum_{g(i) = j} (x_i - \bar{X}_j)(x_i - \bar{X}_j)^T \quad (26)$$

and

$$\overline{W} = \frac{1}{N} \sum_{j=1}^{J} N_j \overline{W_j} \quad (27)$$

The two-class LDA method chooses a single projection that maximizes the ratio of the overall SSQP to the within class SSQP

$$\overrightarrow{\theta_1} = \arg max_{\vec{\theta}_1} \left(\frac{\vec{\theta}_1^T \bar{T} \vec{\theta}_1}{\vec{\theta}_1^T \overline{W} \vec{\theta}_1} \right) \quad (28)$$

It can be shown that the solution to the above equation corresponds to that right eigenvector of $\overline{W}^{-1}\bar{T}$ which has the largest eigenvalue.[9] If the number of classes is more that two, we can find a p dimensional linear transformation *(p < n)* for classification. To get a p-dimensional transformation, we maximize the ratio

$$\hat{\theta}_p = \arg max_{\theta_p} \frac{\left|\theta_p^T \overline{T} \theta_p\right|}{\left|\theta_p^T \overline{W} \theta_p\right|} \qquad (29)$$

To obtain $\hat{\theta}_p$, we choose those eigenvectors $\overline{W}^{-1}\overline{T}$ of that correspond to the largest p eigenvalues, and let $\hat{\theta}_p$ be a $n \times p$ matrix of these eigenvectors. [4][6][10] The p dimensional features thus obtained $y = \hat{\theta}_p^T x$ are uncorrelated. [3]

III. DATA FEATURES AND THEIR PROBLEMS

The data features used in this project are a set of 42 features computed as part of the Scenic Beauty Estimation Project. The features calculated in the Scenic Beauty Estimation Project are explained as follows. Since color is a highly noticeable feature in a forest environment. It is affected by the temporal rhythm of season. Variation of color by season is a highly notable change in forest vegetation. This affects performance of beauty as related to seasonal patterns.[1] The approach here was to extract the mean of each of the color in the image. It was observed from the subjective ratings that summer and spring are preferred over winter. This is based on the fact that people prefer green and blue color as compared to red and yellow which comes from the inclination of the people towards natural colors.[1] Each pixel in PPM is represented by three bytes one byte for each of the color of red, green and blue. Both the number of long lines and short lines were computed as features. The sharpness of the image in each of the colors was computed yielding sharpness of red, sharpness of blue, and sharpness of green.[1] The standard deviation of each color add three more features as standard deviation of red, standard deviation of green, and the standard deviation of blue. Entropy of red, entropy of blue, and entropy of green provided three more features. The final feature used was the compression ratio of the image.[1]

One of the problems in data classification problems using features is the addition of features does not always provide new information for the system. Ideally, the more features take from a data source, more information is gained. Unfortunately, the addition of new features does not necessarily provide more information. [5][6] One of the reasons for this is that in a large collection of features some of the features may be linear combinations or repetitions of features with different weights. This has shown itself to be a problem in the problem of determining a scenic beauty estimation for images.

IV. 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 of reducing the number of features to the most significant features and determining a better estimation of the images scenic beauty. 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 general flow of the algorithm is shown in Figure 4.

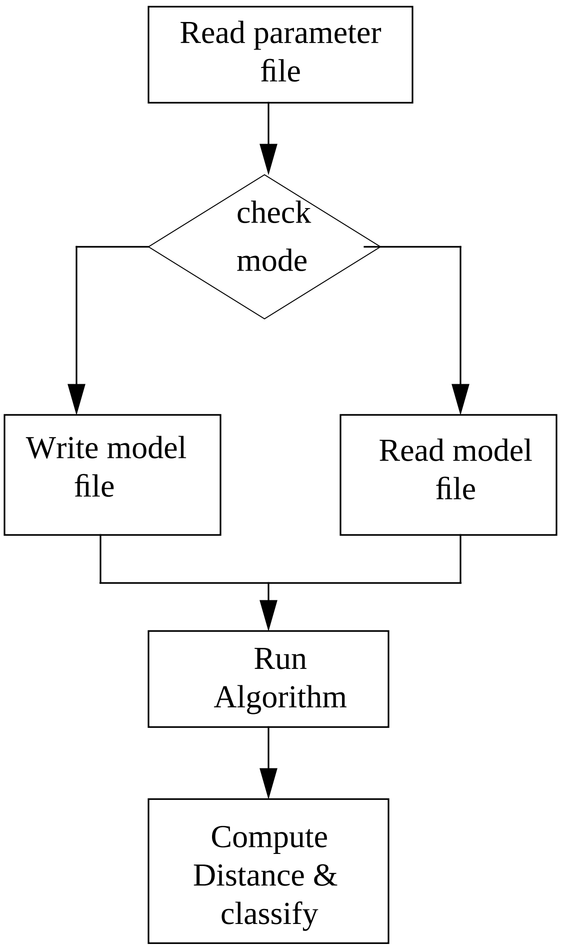

Figure 4: Algorithm flow for the Scenic Beauty Estimation Project.

The algorithm was written dynamically so that any number of features can be used to better enhance the performance ratings. This feature of the software allows any flexibility in the use of large numbers of features to just a few features along with any number of feature reduction for better classification of images. LDA portion of the program allows the user to specify the number of features used in computing the models and the number of features to reduce the feature space to.

All of these features makes the use of this algorithm a great asset.

V. ALGORITHM TESTING

The algorithm for LDA was initially coded in Matlab and tested with the training and testing data consisting of all the features available at that time. Results were obtained by reducing the number of features and computing results for each number of features used. These results were compared to the previous test results and human SBE ratings. These tests provided a new feature space with reduced features where adequate discrimination information is available. The use of six features provided the best performance with an error of 60.25%.

With an initial error of 60.25%, the question of whether the algorithm was operating properly or not arose. To determine if the algorithm was operating properly, the use of synthetic data sets was employed. The first synthetic data set (Figure 5) is comprised of a two feature space (two dimensional data: x and y) with three classes. Each class is a definite cluster of points separated by a distance to allow definite classification by the use of our LDA algorithm. The second synthetic data set, see Figure 6, is comprised of a two feature space with three classes that are overlapping. This data set shows each class of data being overlapped or contained in another class, thus showing little discrimination between each class.

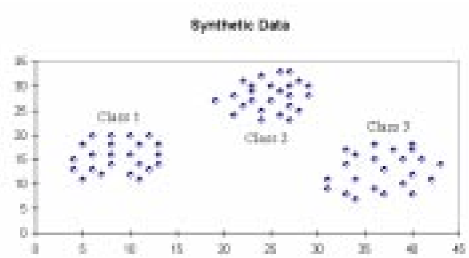

Figure 5: Synthetic test set with three separate and distinct classes.

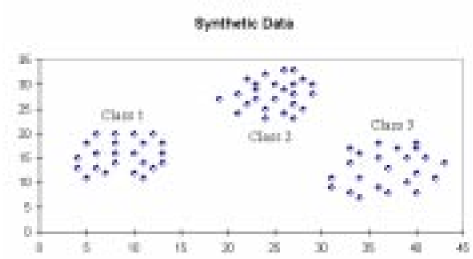

Figure 6: Synthetic test set with three overlapping classes.

The first synthetic data set was used to show that our implementation of LDA would be able to classify data successfully. The second synthetic data set affirmed that LDA would not be able to accurately classify data. These tests helped to confirm that our algorithm was indeed operating properly thus giving confidence in its operation using real data.

By initially developing and testing the LDA algorithm in Matlab, we ensured that the algorithm was working properly before the C++ coding was implemented. This also allowed comparison testing with the Matlab code and the C++ code for debugging and testing the final software.

By developing our algorithm both in Matlab and C++, we were able to compare the final solutions of both algorithms on the test data and the Scenic Beauty Estimation image data. Using a comparison of the results computed using the Matlab version and the C++ versions of our LDA algorithm, we could show that each algorithm is classifying the images properly.

VI. Confirmation of Poor Data

When the algorithm was run on the real data sets, the best error performance obtained was 60.25%. Due to the high rate of error, it was hypothesized that the features being used to classify the images did not provide enough class discrimination to allow accurate classification. To help determine if the data did provide any class discrimination, the algorithm was performed on the features Long Lines and Entropy. The error obtained using these two features was 64.3%. The training data was plotted for Long Lines verse Entropy for each class to help show that these features do not provide adequate discrimination between class -- Figure 7, Figure 8, and Figure 9. It can be seen since each class completely overlaps the other classes that there is no discrimination between each class. It is expected that similar results would occur if other features were plotted in a similar fashion.

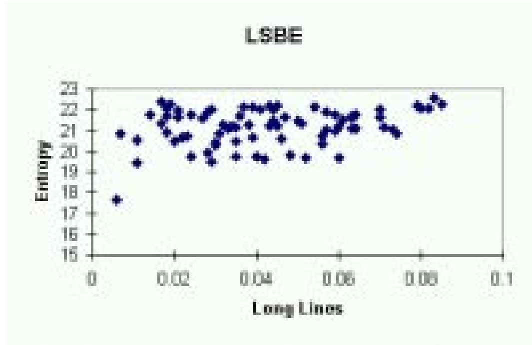

Figure 7: LSBE data for Long Lines and Entropy.

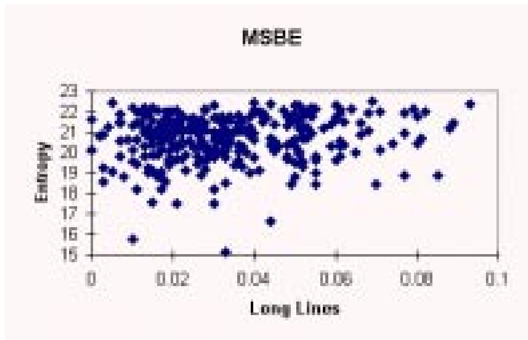

Figure 8: MSBE data for Long Lines and Entropy.

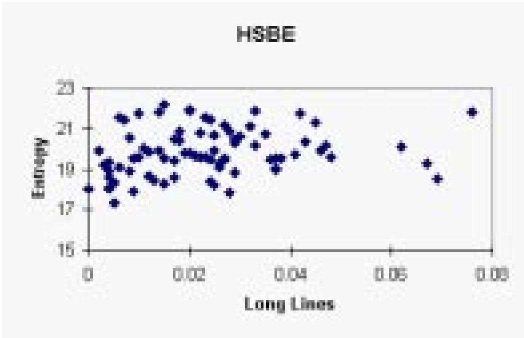

Figure 9: HSBE data for Long Lines and Entropy.

This is a highly significant conclusion concerning the feature data. It shows that there is very little discrimination between classes using the existing features. This would justify the poor performance of LDA on the existing features. If features that contain adequate class discrimination could be obtained, LDA would provide better results for classification.

VII. RESULTS

It has been determined that the use of the LDA algorithm has an error rate of 60.25% for its best performance, which is using six features. This error rate is slightly worse than that of other techniques used in the classification of forest images. Figure 10 shows a comparison for the best performances of the existing algorithms and guessing systems. One interesting problem in the data classification, as mentioned earlier, is the problem of finding the best possible combination of feature data for classification. Figure 11 shows how LDA performed using a range of one to twelve features. It can be seen that LDA had its best performance when using five and six features.

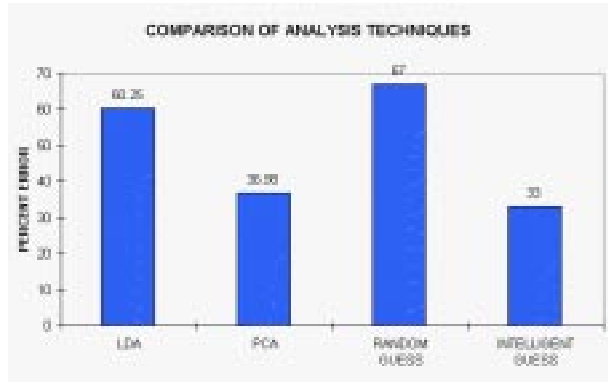

Figure 9: Comparison of analysis techniques.

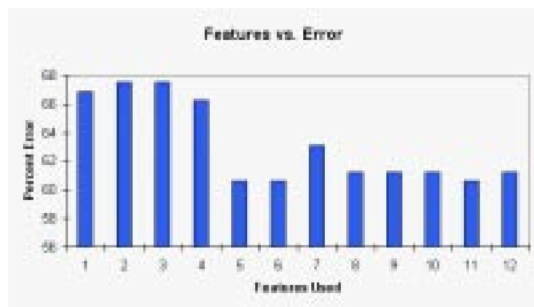

Figure 10: Performance for varying feature reduction.

A comparison of the existing ratings for each image, the Scenic Beauty Rating (SBE), and the Normalized Scenic Beauty Rating (SBEZ) estimations developed using other algorithms, and the results of applying our LDA algorithm are shown with the use of histograms to provide a better understanding of how the LDA classification compares with other classification techniques. The evaluation of this project consisted of comparing the values of the analysis to those standards mentioned above with the use of histograms. With this, it was possible to determine the error between the human SBE rating and the LDA algorithms and to find the difference between the LDA algorithm and the other algorithms existing results.

The matrix listed below is a confusion matrix which illustrates a brief comparison between the human judgement and estimated rating. The results shown are for dimensionality reduction with six features which produced least error performance of 60.25%

	subjective rating			
	lsbe	msbe	hsbe	measured rating
lsbe	19	52	7	
msbe	3	27	4	
hsbe	7	24	17	

VIII. CONCLUSION

The use of Linear Discriminant Analysis has helped to provide better insight to the project of Scenic Beauty Estimation. It has been found that the use of LDA does not provide highly accurate method for classification of the database images. Through the use of comparing the features for Long Lines and Entropy and testing with synthetic data, we have come to the conclusion that the features presently used do not provide enough class discrimination for LDA to provide accurate results.

As discussed earlier, one of the problems using of large numbers of features is encountering features that provide no new information for classification. A problem was encountered with this in the algorithm testing. With the use of 42 features, the within class covariance and the between class covariance matrices produced singular or ill-conditioned matrices. This caused problems when taking the inverse of these matrices in the calculation of the transformation function, due to the fact that the inverse of a singular matrix does not exist. This problem might be avoided by providing a check in the algorithm when the eigenvalues and eigenvectors are being computed to force the values of these matrices from becoming ill-conditioned.

The problem of the non-discriminating features used in the Scenic Beauty Estimation Project is very significant. This has shown that the main problem with the use of LDA for classification of forest images is the limitation of finding features that can show discrimination between the classes. This problem will be encountered in any problem where LDA is being used. It essentially comes down to the question of does the problem being solved fit into the model that LDA performs well in. In the case of the Scenic Beauty Estimation Project, the present features do not allow fit the model of LDA well enough to obtain accurate results.

IX. FUTURE RESEARCH DIRECTIONS

We envisage our future research efforts to be directed towards the development of more features of the images in the database which could offer more discrimination among the classes. The continuation of implementing more algorithms for data classification could also help to improve and forward the Scenic Beauty Estimation Project. Apart from this, a graphical user interface tool to be developed which offers the user various options of selecting the required number of features, training the data as well testing any required image.

X. ACKNOWLEDGEMENTS

We wish to thank Nirmala Kaldinidi for her numerous contributions in the development of software and database, and her continued support throughout this project. We also wish to acknowledge the continual support and guidance we have received from Dr.Joe Picone and the ISIP (Institute of Signal and Information Processing) group.

XI. REFERENCES

[1] Nirmala Kalidindi, Audrey Le, Dr. Joe Picone, Dr. Victor A. Rudis, Scenic Beauty Estimation, Research Paper, Mississippi State University, 1996.

[2] Nagendra Kumar, Doctoral Thesis. John Hopkins University, 1996.

[3] N. Kumar, A. Andreou, “On Generalizations of Linear Discriminant Analysis.” Research Paper, John Hopkins University, 1996.

[4] W. R. Dillon, M. Goldstein, Multivariate Analysis. John Wiley & Sons, Inc., 1984.

[5] G. F. Hughes, “One the Mean Accuracy of Statistical Pattern Recognizer.” IEEE Transactions on Information Theory, 14(1):55-63.

[6] L. Kanal, B. Chandrasekaran, “On Dimensionality and Sample Size in Statistical Pattern Classification.” Pattern Recognition, 1971.

[6] Harry R. Barker, Barbara M. Barker, Multivariate Analysis of Variance (MANOVA). The University of Alabama Press, 1984.

[7] Richard O. Duda, Peter E. Hart, Pattern Classification and Scene Analysis. John Wiley & Sons, Inc., 1973.

[8] Shaffer, Ron, Discriminant Analysis For Chemistry Problems. Naval Research Laboratory.

[9] R. A. Fisher, The Use of Multiple Measurements in Taxonomic Problems. Ann. Eugen.,1936.

[10] C. R. Rao, Linear Statistical Inference and Its Application. John Wiley and Sons, New York 1965.

[11] Duane Hanselman, Bruce Littlefield, The Student Edition of MATLAB Version 5 User’s Guide. Prentice-Hall, Inc., 1997.

[12] R. J. Schalkoff, “Digital Image Processing and Computer Vision, “John Willey & Sons, Inc., 1989.

[13] K. R. Castleman, Digital Image Processing, John Willey & Sons, Inc., 1978

[14] W.K.Pratt, Digital Image Processing. John Willey & Sons, Inc., 1978

[15] William H.Press, Brian P.Flannery, Saul A.Teukolsky, William T.Vetterling, Numerical Recipes in C, The Art of Scientific Computing.