COMPARATIVE STUDY OF DIMENSIONALITY REDUCTION TECHNIQUES USING PCA AND LDA FOR CONTENT BASED IMAGE RETRIEVAL

Shereena V. B.[1] and Julie M. David[2]

[1,2] Asst. Professor, Dept. of Computer Applications, MES College, Marampally, Aluva, Cochin, India

[1]shereenavb@gmail.com

[2]julieeldhosem@yahoo.com

ABSTRACT

The aim of this paper is to present a comparative study of two linear dimension reduction methods namely PCA (Principal Component Analysis) and LDA (Linear Discriminant Analysis). The main idea of PCA is to transform the high dimensional input space onto the feature space where the maximal variance is displayed. The feature selection in traditional LDA is obtained by maximizing the difference between classes and minimizing the distance within classes. PCA finds the axes with maximum variance for the whole data set where LDA tries to find the axes for best class seperability. The proposed method is experimented over a general image database using Matlab. The performance of these systems has been evaluated by Precision and Recall measures. Experimental results show that PCA based dimension reduction method gives the better performance in terms of higher precision and recall values with lesser computational complexity than the LDA based method.

KEYWORDS

Color histogram, Feature Extraction, Euclidean distance, Principal Component Analysis, Linear Discriminant Analysis, Eigen Values, Eigen Vectors.

1. INTRODUCTION

As we know human beings are predominantly visual creatures. The visualisation of the images which we see, in real or imaginary, make sense of the world around us to identify and differentiate the things which we see at a quick glance. We are bestowed with very precise visual skills to identify an image by size and also by differentiating the colors. We can process a large amount of visual information very quickly.

An image processing task consists of acquiring the image, pre-processing, segmentation, representation and description and finally recognition and interpretation. There are four types of digital images, binary, grey scale, true color or RGB and indexed [1]. Binary representation images include text, fingerprints or architectural plans where each pixel is black or white. Grey

Dhinaharan Nagamalai et al. (Eds) : CSTY, SIGI, AI, MaVaS - 2015
pp. 41–55, 2015. © CS & IT-CSCP 2015 DOI : 10.5121/csit.2015.50905

scale images consist of X-rays, images of printed works etc where each pixel is a shade of grey, normally from 0 to 255. True color or RGB images are the color images where each pixel is described by the amount of red, green and blue in it. Finally there are indexed images where the image has an associated color map which is a list of all the colors used in that image. Each pixel has a value which does not give its color, but an index to the color in the map.

There has been a tremendous growth in the digital information over years. This trend has motivated research in image databases, which were nearly ignored by traditional computer systems due to the enormous amount of data necessary to represent images and the difficulty of automatically analyzing images. Currently, storage is less of an issue since huge storage capacity is available at low cost. Large image databases are used in many application areas such as satellite imaging, and biometric databases, Crime prevention, military, Intellectual property, Architectural and engineering design, Fashion and interior design, Journalism and advertising, Medical diagnosis, Geographical information and remote sensing systems, Cultural heritage, Education and training, Home entertainment, Web searching, where it is important to maintain a high degree of precision [2]. Thus an important issue was the fast image retrieval from large databases. This trend led to the development of research area known as Content Based Image Retrieval. CBIR systems retrieves features from the raw images themselves and calculate an association measure between the query image and database images based on these features. We need to develop an efficient system for retrieving images since speed and precision are important.

CBIR consists of different stages such as Image acquisition, image Pre-Processing, Feature Extraction, Similarity Matching and obtain the resultant images. Image Acquisition is the process of acquiring a digital image database which consists of n number of images. The Pre-processing stage involves filtering, normalization, segmentation, and object identification. The output of this stage is a set of significant regions and objects. In the Feature extraction stage, visual information such as color and texture is extracted from the images and saves them as feature vectors in a feature vector database. One of the major problems with Content Based image retrieval system is the large number of features extracted which requires large amount of memory and computation power. To overcome this problem we have to construct a combination of features which best describe the data with sufficient accuracy. So in this stage, we use dimension reduction algorithms which extract only essential features from the feature vector database and store them as reduced feature vector database. Thus the output of feature extraction stage is a reduced set of features which best describes the image. In the Similarity matching stage, the reduced feature vectors of query image calculated is matched with the feature vectors of reduced feature vector database using any of the Distance methods available such as Euclidean distance, City Block Distance, Canberra Distance [3].

The most popular among the Dimensionality Reduction Algorithms are Principal Component Analysis and Linear Discriminant Analysis. Principal Component Analysis defines new attributes (principal components or PCs) as mutually-orthogonal linear combinations of the original attributes. For many image datasets, it is sufficient to consider only the first few PCs, thus reducing the dimension. Linear Discriminant Analysis [4] easily handles the case where the within-class frequencies are unequal and their performances have been examined on randomly generated test data. This method maximizes the ratio of between-class variance to the within-class variance in any particular data set thereby guaranteeing maximal separability. In this paper, we compare the above two dimensionality reduction techniques by implementing the algorithms on a given image data set. Experimental results on database images shows that the feature set can be considerably reduced without significant degradation in performance.

The rest of this paper is organized as follows. Section 2 deals with Literature Review. In Section 3, we explain Proposed Methodology. Section 4 consists of Comparative study of PCA and LDA, Conclusions are given in Section 5.

2. LITERATURE REVIEW

H.H. Pavan Kumar Bhuravarjula and VNS Vijayakumar proposed in their paper "A novel content based image retrieval using variance color moment" that color moments gives average high precision and recall [2]. In the paper of Manimala Singha and K. Hemachandran [5], they presented a novel approach for Content Based Image Retrieval by combining the color and texture features called Wavelet-Based Color Histogram Image Retrieval (WBCHIR). The experimental result shows that the proposed method outperforms the other retrieval methods in terms of Average Precision. Pranali Prakash Lokhande , P. A. Tijare [6] concluded in their paper " Feature Extraction Approach for Content Based Image Retrieval "that the combination of the color and texture features of an image in conjunction with the shape features will provide a robust feature set for image retrieval. S. Mangijao Singh and K. Hemachandran [7] in their paper "Content-Based Image Retrieval using Color Moment and Gabor Texture Feature" proposed an efficient image retrieval method based on color moments and Gabor texture features. To improve the discriminating power of color indexing techniques, they encoded a minimal amount of spatial information in the index. Mohd. Danish, Ritika Rawat, Ratika Sharma [3] in their paper "A Survey: Content Based Image Retrieval Based On Color, Texture, Shape and Neuro Fuzzy" provides an overview of the functionality of content based image retrieval systems. Most systems use color and texture features, and some systems use shape features.

A. Ramesh Kumar and D. Saravanan in their paper "Content Based Image Retrieval Using Color Histogram" [8], CBIR using color histograms technique is proposed with help of principal component analysis technique to improve the image retrieval performance. Swati V. Sakhare and Vrushali G. Nasre, [9] in their paper "Design of Feature Extraction in Content Based Image Retrieval (CBIR) using Color and Texture" designed an application which performs a simple color-based search in an image database for an input query image, using color, texture and shape to give the images which are similar to the input image as the output. The number of search results may vary depending on the number of similar images in the database. In the paper "A Proposed Method for Image Retrieval using Histogram values and Texture Descriptor Analysis" [10], Wasim Khan, Shiv Kumar. Neetesh Gupta and Nilofar Khan proposed a method for image retrieval using histogram values and texture descriptor analysis of image. When a query image is submitted, its color and texture value is compared with the color and texture value of different images stored in database. The images having closest value compared to query image are retrieved from database are displayed on GUI as result.

S. Meenachi Sundaresan and Dr. K.G. Srinivasagan [11] proposed in their paper "Design of Image Retrieval Efficacy System Based on CBIR" that the performance of a retrieval system can be measured in terms of its recall (or sensitivity) and precision (or specificity). Recall measures the ability of the system to retrieve all models that are relevant, while precision measures the ability of the system to retrieve only models that are relevant. In the paper " An Enhancement on Content-Based Image Retrieval using Color and Texture Features", [12] Tamer Mehyar, Jalal Omer Atoum proposed an enhancement on the use of color and texture visual features in Content-Based Image Retrieval (CBIR) by adding a new color feature called Average Color Dominance which tries to enhance color description using the dominant colors of an image.

In the paper "Implementation of Principal Component Analysis with Fuzzy Annotation for CAD Jewellery Images", Pinderjeet Kaur [13] proposed that Principal Component Analysis (PCA) can be used for dimension reduction to reduce the computation cost for the system of Content Based Image Retrieval (CBIR). Arunasakthi. K, KamatchiPriya. L [14] stated in their paper "A Review On Linear And Non-Linear Dimensionality Reduction Techniques" that Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) are regarded as the most fundamental and powerful tools of dimensionality reduction for extracting effective features of high-dimensional vectors in input data. According to Julie M. David and Kannan Balakrishanan, principal components are new set of variables which are generated by the application of dimensionality reduction method [15]. The basic procedures behind PCA are (i) the inputs data are normalized, so that each attribute falls within the same range. This helps ensure that attributes with large domains will not dominate attributes with smaller domains, (ii) PCA computes k orthonormal vectors that provides a basis for the normalized input data. These are unit vectors that each point in a direction perpendicular to the others. These vectors are referred to as the principal components and (iii) The principal components are sorted in order of decreasing strength. [16]

Kresimir Delac, Mislav Grgic and Sonja Grgic [17] in their paper "Independent Comparative Study of PCA, ICA, and LDA on the FERET Data Set" proposed that PCA finds a set of the most representative projection vectors such that the projected samples retain most information about original samples whereas LDA uses the class information and finds a set of vectors that maximize the between-class scatter while minimizing the within-class scatter. In the paper "CBIR Feature Vector Dimension Reduction with Eigenvectors of Covariance Matrix using Row, Column and Diagonal Mean Sequences" [18], Dr. H.B. Kekre, Sudeep D. Thepade and Akshay Maloo stated that PCA can be used to transform each original image from database into its corresponding eigen image.

In the paper "Linear Discriminant Analysis bit by bit" Sebastian Raschka [19] stated that PCA can be described as an unsupervised algorithm, since it ignores class labels and its goal is to find the directions (the so-called principal components) that maximize the variance in a dataset. In contrast to PCA, LDA is supervised and computes the linear discriminants that will represent the axes that maximize the separation between multiple classes.

The main motivation of this work is to compare two dimension reduction techniques PCA and LDA to find out which of them selects the best features from the feature set to reduce the dimensions of the dataset with minimal loss of information. Principal Component Analysis (PCA) is a mathematical tool used to extract principal components of original image data. These principal components may also be referred as Eigen images. Linear Discriminant Analysis seeks to reduce dimensionality while preserving as much of the class discriminatory information as possible. In LDA, we compute eigenvectors from our dataset and collect them in scatter matrices.

3. PROPOSED METHODOLOGY

3.1 Prepare input data.

In this paper, a general image database consisting of 500 images is used for testing the comparative study of PCA and LDA. Principal Component Analysis defines new attributes as mutually-orthogonal linear combinations of the original attributes. Linear Discriminant Analysis computes the linear discriminants that will represent the axes that maximize the separation between multiple classes. In order to obtain better search results and to express more image information, we consider the dominant color and texture features combined. These low level features are extracted using color moments, color histogram, color auto correlogram and wavelet. The basis of color moments is that the distribution of color in an image can be considered as a probability distribution which can be characterized by various moments [20]. The color histogram for an image is constructed by quantizing the colors within the image and counting the number of pixels of each color. The color correlogram was proposed to characterize not only the color distributions of pixels, but also the spatial correlation of pairs of colors. Wavelet Analysis is a popular method for extracting texture from an image. The discreet wavelet transform (DWT) of a signal is calculated by passing it through a series of filters (high and low pass filters) and then down-sampled [21].

3.1.1 Color Moments

The first order color moment(Mean), Second order color moment(Standard deviation) and the third order color moment(Skewness) have been used for color feature extraction [20]. Since only 9 (three moments for each of the three color components R, G, B) numbers are used to represent the color content of each image, color moments are a very compact representation compared to other color features.

$$\mu_i = \frac{1}{N}\sum_{j=1}^{n} P_{ij}$$

$$\sigma_i = (\frac{1}{N}\sum_{j=1}^{n}((P_{ij} - \mu_i)^2)^{\frac{1}{2}}$$

$$S_i = (\frac{1}{N}\sum_{j=1}^{n}((P_{ij} - \mu_i)^3)^{\frac{1}{3}}$$

Where P_{ij} is the value of the i- the color channel of image pixel j and N is the number of pixels in the image.

When a query image is submitted for image retrieval, its color moments are extracted and added to feature set for matching the image with the images stored in the database. The following are the steps for extracting color moments from an image.

1. Extract the values of each plane R, G, B corresponding to the image.
2. Find the mean , standard deviation and skewness of each plane
3. Convert to column vector output of the moments.

The following table gives the color moments of 5 images where M corresponds to mean, Std corresponds to standard deviation, Skew corresponds to Skewness and R for Red, G for Green and B for Blue plane respectively.

Table 1. Color Moments Table

M(R)	Std(R)	Skew(R)	M(G)	Std(G)	Skew(G)	M(B)	Std(B)	Skew(B)
0.4372	0.3659	0.2132	0.1925	0.1821	0.3083	0.0014	0.0013	0.0009
0.4385	0.3409	0.2389	0.2018	0.1928	0.3131	0.0011	0.0015	0.0010
0.4351	0.3572	0.2582	0.2521	0.2349	0.3069	0.0019	0.0033	0.0030
0.5061	0.4364	0.2362	0.2355	0.2283	0.4019	0.0008	0.0015	0.0011
0.3765	0.4012	0.2818	0.2850	0.2844	0.3319	0.0054	0.0064	0.0046

3.1.2 Color Histogram

A histogram is a graph that represents all the colors and the level of their occurrence in an image irrespective of the type of the image [8]. This technique describes the proportion of pixels of each color in an image. It has been used as one of the feature extraction attributes with the advantage like robustness with respect to geometric changes of the objects in the image. The color histogram is obtained by quantizing image colors into discrete levels and then counting the number of times each discrete color occurs in the image. In a CBIR system, a query image is compared with the histograms of all the images in database [22].

A color histogram H for a given image is defined as a vector

$$H = \{ H[1], H[2], \ldots H[i], \ldots, H[N]\}$$

where i represent a color in the color histogram, H[i] is the number of pixels in color i in that image, and N is the number of bins in the color histogram, i.e., the number of colors in the adopted color model.

In order to compare images of different sizes, color histograms should be normalized. The normalized color histogram H ′ is defined as

$$H' = \{H'[0], H'[1], \ldots H'[i], \ldots H'[N]\}$$

where $H'[i] = \frac{H[i]}{XY}$, XY is the total number of pixels in an image.

From the query image submitted for image retrieval, its color histogram features are extracted and added to feature set for matching the image with database images. The following steps give a method to calculate color histogram.

1. Convert the image from RGB color space to HSV color space.
2. Define number of clusters for each HSV plane.
3. Find the maximum value of each plane.
4. Cluster each values after normalisation.
5. Add each color to any one of the appropriate cluster.
6. Find the probabilistic values and convert the values to the column vector.

3.1.3 Color autocorrelogram

A color correlogram is a table indexed by color pairs, where the k-th entry for (i, j) specifies the probability of finding a pixel of color j at a distance k from a pixel of color i in the image [20]. Let I represent the entire set of image pixels and $I_{c(i)}$ represent the set of pixels whose colors are c(i).Then, the color correlogram is defined as:

$$\gamma_{(i,j)}(k) = Pr_{p1 \in c(I), p2 \in I} \ [p2 \in I_{c(j)} | p1 - p2| = k]$$

Where i, j ∈ {1, 2, …, N}, k∈ {1, 2, …, d}, and | p1 – p2 | is the distance between pixels p1 and p2.

The color auto correlogram of the query image is extracted and added to feature vector for the extraction of similar database images. The following are the steps for extracting correlogram features from an image.

1. Reduce the number of colors in the RGB image.
2. Correlate each pixel with the neighbourhood pixels for getting the correlogram vector.

3.1.4 Texture

Like color, the texture is a powerful low-level feature for image search and retrieval applications. The texture measures try to retrieve the image or image parts characteristics with reference to the changes in certain directions and the scale of the images. This is most useful for images with homogeneous texture [3]. Wavelet analysis is an exciting new method for solving difficult problems in mathematics, physics, and engineering, with modern applications as wave propagation, data compression, signal processing, image processing, pattern recognition, computer graphics, the detection of aircraft and submarines and other medical image technology. A wavelet is a mathematical function used to divide a given function into different frequency components [21]. A wavelet transform is the representation of a function by wavelets, which represent scaled and translated copies of a finite length or fast-decaying oscillating waveform (known as the "mother wavelet"). The Wavelet transform of a function is the improved version of Fourier transform. Wavelet transforms have advantages over traditional Fourier transforms because local features can be described better with wavelets that have local extent. Some mother wavelet families implemented in Matlab are Daubechies, Symlet, Coiflet, Biortogonal and Reverse biorthogonal wavelets) and the fractional B-spline functions are used to compute different feature vectors. Orthogonal wavelets with FIR filters can be defined through a scaling filter. Predefined families of such wavelets include Haar, Daubechies, Symlets and Coiflets. In this paper, Coiflet wavelet function is used to extract texture features. The following steps give a method to calculate Texture of an image.

1. Convert the image to grayscale.
2. Find the 4 stage Coif wavelet coefficients.
3. Find the mean and standard deviation of the above coefficients and output to a column vector

The following table gives the 4 stage coiflet texture values of 5 images.

Table 2. Coiflet Texture values Table

5.9048	2.6054	0.1637	0.0979
5.5827	2.9509	0.2472	0.0167
6.2997	3.7201	0.2554	0.0405
7.1697	3.6297	0.2840	0.0528
6.5487	4.3823	0.0779	0.2445

The following table gives the first 10 features of 5 images in the database before applying dimension reduction algorithms.

Table3. Table of features before Dimension Reduction

1	2	3	4	5	6	7	8	9	10
0.4372	0.3659	0.2132	0.1925	0.1821	0.3083	0.0014	0.0013	0.0009	5.9048
0.4385	0.3409	0.2389	0.2018	0.1928	0.3131	0.0011	0.0015	0.0010	5.5827
0.4351	0.3572	0.2582	0.2521	0.2349	0.3069	0.0019	0.0033	0.0030	6.2997
0.5061	0.4364	0.2362	0.2355	0.2283	0.4019	0.0008	0.0015	0.0011	7.1697
0.3765	0.4012	0.2818	0.2850	0.2844	0.3319	0.0054	0.0064	0.0046	6.5487

3.2 Principal Component Analysis (PCA) Vs Linear Discriminant Analysis (LDA)

Principal Component Analysis is a technique which uses sophisticated underlying mathematical principles to transform a number of possibly correlated variables into a smaller number of variables called principal components [13]. It is one of the most important results from applied linear algebra. The advantage of PCA is finding the patterns in the data and compressing data by reducing the number of dimensions without loss of information. The mathematical concepts that are used for PCA are Standard Deviation, Variance, Co–variance and Eigenvectors [23]. The database images belonging to same category may differ in lighting conditions, noise etc., but are not completely random and in spite of their differences there may present some patterns. Such patterns could be referred as principal components. PCA is a mathematical tool used to extract principal components of original image data. These principal components may also be referred as Eigen images [18]. An important feature of PCA is that any original image from the image database can be reconstructed by combining the eigen images. The algorithm to calculate Principal Components is as follows.

1. Represent the image as one dimensional vector of size N x N.
   Suppose we have M vectors of size N (= rows of image × columns of image) representing a set of sampled images. Then the training set becomes: Γ1, Γ2, Γ3.....ΓM.

2. The Mean value of the pixels intensities in each image is calculated and subtracted from the corresponding image. The process is continued for all images in the database.

3. The covariance matrix which is of the order N^2 x N^2 is calculated as given by $C = AA^T$.

4. Find the Eigen values of the covariance matrix C by solving the equation $(C\lambda - I) = 0$ .To find the eigenvector X repeat the procedure where Xi indicates corresponding Eigen values.

5. The Eigen vectors are sorted according to the corresponding Eigen values in descending order.

6. Choose the First ‘K’ Eigen vectors and Eigen Values.

The following table gives the first 10 features of 5 images after applying the dimension reduction algorithm PCA and reducing the feature database.

Table 4. Table of features after Dimension Reduction

1	2	3	4	5	6	7	8	9	10
62.189	7.105e-15	-1.776e-15	1.154e-14	4.024e-15	-1.318e-15	-2.088e-15	-3.344e-15	9.992e-16	-1.089e-15
63.286	-2.442e-14	2.220e-15	3.996e-15	-1.554e-15	-3.885e-16	-2.396e-16	-1.707e-15	1.498e-15	9.436e-16
60.849	-3.497e-15	-1.065e-14	7.771e-16	2.220e-15	-2.657e-15	-5.568e-16	1.020e-15	-2.636e-15	3.677e-16
64.930	-1.776e-14	-5.329e-15	4.218e-15	1.110e-16	-1.845e-15	7.216e-16	-3.486e-16	-2.713e-15	-3.747e-15
51.539	8.882e-16	8.881e-15	-7.549e-15	-4.218e-15	5.343e-16	-2.331e-15	-2.307e-15	1.332e-15	1.637e-15

Linear Discriminant Analysis (LDA) [24] is most commonly used as dimensionality reduction technique in the pre-processing step for pattern-classification and machine learning applications. The feature selection in traditional LDA [14] is obtained by maximizing the difference between classes and minimizing the distance within classes. LDA finds the vectors in the underlying space that best discriminate among classes. The prime difference between LDA and PCA is that PCA does more of feature classification and LDA does data classification [4].

1. Compute the *d*-dimensional mean vectors for the different classes from the dataset.
2. Compute the scatter matrices (between-class and within-class scatter matrix).
3. Compute the eigenvectors ($\mathbf{e_1}, \mathbf{e_2}, ..., \mathbf{e_d}$) and corresponding eigen values ($\boldsymbol{\lambda_1}, \boldsymbol{\lambda_2}, ..., \boldsymbol{\lambda_d}$) for the scatter matrices.
4. Sort the eigenvectors by decreasing eigenvalues and choose **k** eigenvectors with the largest eigenvalues to form a $d \times k$-dimensional matrix **W** (where every column represents an eigenvector).
5. Use this $d \times k$ eigenvector matrix to transform the samples onto the new subspace. This can be summarized by the mathematical equation: $\mathbf{y} = \mathbf{W}^T \times \mathbf{x}$ (where **x** is a $d \times 1$-dimensional vector representing one sample, and **y** is the transformed $k \times 1$-dimensional sample in the new subspace).

3.3 Similarity Matching

If R' be the dimensionality reduced feature database and R" is the feature vector obtained from query image, then the retrieval system is based on a similarity measure defined between R' and R" [25]. In this paper, Euclidean distance is used to measure the similarity between the feature vectors of reduced query image and reduced database images. The formula for Euclidean distance [26] is given as

$$Euclidean\ Distance = \sqrt{\sum_{i=1}^{n}(Q_i - D_i)^2}$$

Where Q and D are feature vectors of the Query image and database image. After finding the Euclidean Distance, the distances are sorted and the top six images closer to the query image are retrieved.

3.4 Performance Evaluation

The performance of retrieval of the system can be measured in terms of its Recall and Precision. Recall measures the ability of the system to retrieve all the models that are relevant, while Precision measures the ability of the system to retrieve only the models that are relevant [20].

$$Precision = \frac{Number\ of\ relevant\ images\ retrieved}{Total\ Number\ of\ images\ retrieved}$$

$$Recall = \frac{Number\ of\ relevant\ images\ retrieved}{Total\ no\ of\ relevant\ images}$$

The number of relevant items retrieved is the number of the returned images that are similar to the query image in this case. The total number of images retrieved is the total number of images that are returned by the retrieval system. In precision and recall, crossover is the point on the graph where the both precision and recall curves meet. The higher the number of crossover points better will be the performance of the system.

4. COMPARATIVE STUDY OF PCA AND LDA

The proposed method has been implemented using Matlab 13 and tested on a general-purpose database containing 500 images, in JPG format of size 256X384 resized to 286x340. The database includes 500 color images categorized into five classes and each class includes 100 images as follows: African people, Beach, Building, Bus, Dinosaurs. The search is based on the similarity of feature vectors. We have followed the image retrieval technique, as described in the section 3 on different feature extraction schemes such as color and texture. This scheme calculated 110 features by means of histogram, moments, correlogram and Coif wavelet. Further, Principal Component Analysis technique and Linear Discriminant Analysis technique is used to extract the best features from the images. By means of PCA and LDA, the feature set is reduced to 75.Then the reduced query image is compared with the reduced database feature set using Euclidean Distance and the top 6 nearer images are displayed. The quality of the image retrieval, with different feature extraction schemes has been evaluated by randomly selecting query images, of each category, from test image database. Each query returns the top 6 images from database. To measure retrieval effectiveness for the image retrieval system, Precision and Recall values are used. The Precision Recall rates and plots for PCA, LDA and without dimension reduction methods are shown in figure1. The graphical user interface for the retrieval of images using dimension reduction with PCA and LDA are shown in the figure 2 and figure3 respectively. From the GUI, the database is to be selected first using Select Database button, i.e., the database of 500 images. Then the query image is selected from a set of test images using Select Query button. The query can be processed under 3 options- without dimension reduction, Dimension Reduction using PCA and Dimension Reduction using LDA. The images are retrieved based on the option selected and top 6 images are displayed in the Returned images frame. The Precision-Recall Plot gives the Precision and Recall rates of the selected option. The Performance Comparison button shows the Precision-Recall plots of all the three methods of the selected query image. From the Table 5 of Precision and Recall, it is found that the rates are higher for dimension reduction using Principal Component Analysis when compared to Linear Discriminant Analysis. This shows that PCA is a better dimension reduction tool when compared to LDA.

Table 5. Table of Precision Recall Rates

Class	**Without Dimension Reduction**		**LDA**		**PCA**	
	Precision	**Recall**	**Precision**	**Recall**	**Precision**	**Recall**
Tribal	0.6110	0.5	0.6667	0.35	0.667	0.47
Beach	0.8330	0.46	0.5	0.25	0.778	0.37
Towers	0.6666	0.39	0.6667	0.37	0.667	0.37
Bus	0.722	0.4	0.333	0.27	0.667	0.31
Dinosaur	1	0.97	1	0.77	1	0.97

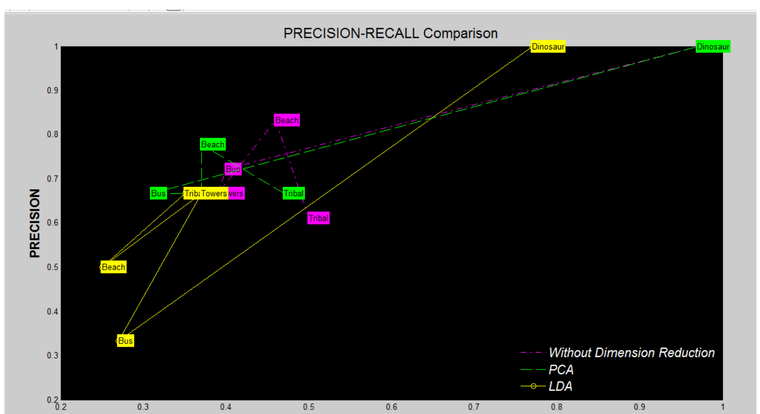

Figure1. Precision –Recall Plot

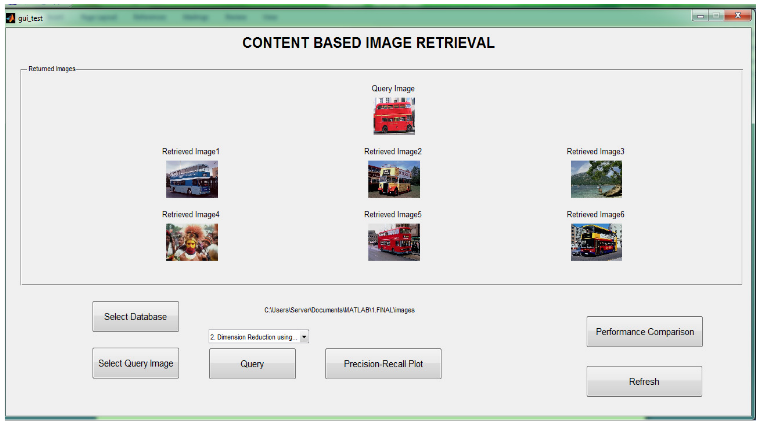

Figure 2.Dimension Reduction using PCA

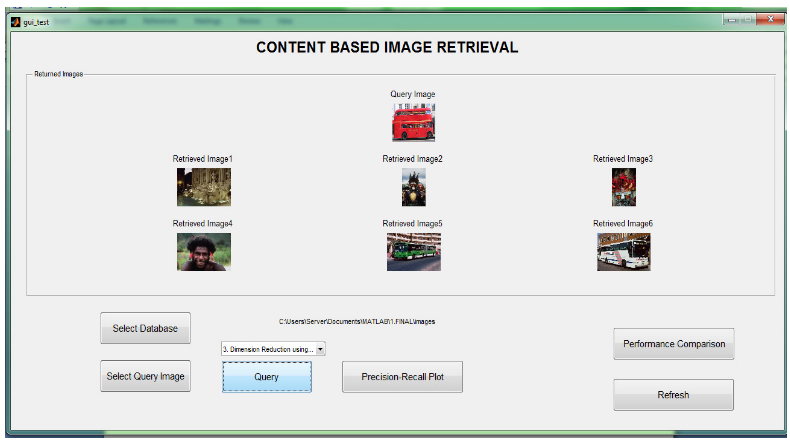

Figure 3. Dimension Reduction using LDA

5. CONCLUSION

In this paper, we presented a comparative study of two dimension reduction methods namely Principal Component Analysis and Linear Discriminant Analysis. Dimensionality reduction methods aim at revealing meaningful structures and unexpected relationships in multivariate data. PCA projects correlated variables into a lower number of uncorrelated variables called principal components. By using only the first few principal components or eigen vectors, PCA makes it possible to reduce the number of significant dimensions of the data, while maintaining the maximum possible variance. The objective of LDA is to perform dimensionality reduction while preserving as much of the class discriminatory information as possible. In LDA, we will compute eigenvectors from our dataset and collect them in scatter-matrices, the between-class scatter matrix and within-class scatter matrix. From the Precision and Recall rates of PCA and LDA calculated in the table in the above section, it can be found that the rates are high for all the cases of PCA when compared to LDA. Thus it is concluded that PCA tends to outperform LDA in almost all cases and hence PCA can be adopted as an effective tool for dimension reduction.

REFERENCES

[1] Rafael Gonzalez and Richard E. Woods, ‘Digital Image processing’, Addison Wesley, 2nd Edn., 2002.

[2] H.H. Pavan Kumar Bhuravarjula And V.N.S Vijaya Kumar(2012), ‘A Novel Content Based Image Retrieval Using Variance Color Moment’, Int. J. Of Computer And Electronics Research,Vol. 1, Issue 3, ISSN: 2278-5795, pp.93-99.

[3] Mohd. Danish, Ritika Rawat and Ratika Sharma(2013) , ‘A Survey: Content Based Image Retrieval Based On Color, Texture, Shape & Neuro Fuzzy’, Int. Journal Of Engineering Research And Applications ISSN : 2248-9622, Vol. 3, Issue 5, Sep-Oct 2013, pp.839-844 .

[4] S. Balakrishnama and A. Ganapathiraju, ‘Linear Discriminant Analysis - A Brief Tutorial’. Institute for Signal and Information Processing, Department of Electrical and Computer Engineering Mississippi State University.

[5] Manimala Singha and K. Hemachandran (2012) ‘Content based image retrieval using color and texture’, Signal & Image Processing : An Int. J. (SIPIJ), Vol.3, No.1, pp.39-57.

[6] Pranali Prakash Lokhande and P. A. Tijare(2012), ‘Feature Extraction Approach for Content Based Image Retrieval’ , Int.Journal of Advanced Research in Computer Science and Software Engineering , Vol. 2, Issue 2, ISSN: 2277 128X .

[7] S. Mangijao Singh , K. Hemachandran (2012) “Content-Based Image Retrieval using Color Moment and Gabor Texture Feature”, Int.J. of Computer Science Issues,Vol. 9, Issue 5, No 1, pp.299-309.

[8] A.Ramesh Kumar and D.Saravanan ,’ Content Based Image Retrieval Using Color Histogram’, Int. J. of Computer Science and Information Technologies, Vol. 4 (2) , 2013, 242 – 245, ISSN:0975-9646.

[9] Swati V. Sakhare & Vrushali G. Nasre(2011), ‘Design of Feature Extraction in Content Based Image Retrieval (CBIR) using Color and Texture’, Int. J. of Computer Science & Informatics, Vol.-I, Issue-II, pp.57-61.

[10] Wasim Khan, Shiv Kumar. Neetesh Gupta and Nilofar Khan(2011), ‘A Proposed Method for Image Retrieval using Histogram values and Texture Descriptor Analysis’, Int. J.of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Vol.-I Issue-II pp.33-36.

[11] S.Meenachi Sundaresan and Dr. K.G.Srinivasagan(2013), ‘Design of Image Retrieval Efficacy System Based on CBIR’, Int. J. of Advanced Research in Computer Science and Software Engineering , Vol. 3, Issue 4, ISSN: 2277 128X, pp.48-53.

[12] Tamer Mehyar, Jalal Omer Atoum(2012), ‘An Enhancement on Content-Based Image Retrieval using Color and Texture Features’, Journal of Emerging Trends in Computing and Information Sciences, Vol. 3, No. 4,ISSN 2079-8407, pp488-496.

[13] Pinderjeet Kaur (2012), ‘Implementation of Principle Component Analysis with Fuzzy Annotation for CAD Jewellery Images’, Int. J. of Emerging Trends & Technology in Computer Science,Vol. 1, Issue 4. ISSN 2278-6856.

[14] Arunasakthi. K and Kamatchipriya. L(2014), ‘A Review On Linear And Non-Linear Dimensionality Reduction Techniques’, Machine Learning And Applications: An Int. J. (Mlaij), Vol.1, No.1, Pp.65-76.

[15] Julie M. David, Kannan Balakrishnan, (2014), Learning Disability Prediction Tool using ANN and ANFIS, International Journal of Soft Computing, Springer Verlag Berlin Heidelberg, ISSN 1432-7643 (online), ISSN 1433-7479 (print), DOI: 10.1007/s00500-013-1129-0, Vol. 18, Issue 6, pp 1093-1112

[16] Julie M. David, Kannan Balakrishnan, (2012), Attribute Reduction and Missing Value Imputing with ANN: Prediction of Learning Disabilities, International Journal of Neural Computing & Applications, Springer-Verlag London Limited, DOI: 10.1007/s00521-011-0619, Vol. 21, Issue 7, pp 1757-1763

[17] Kresimir Delac, Mislav Grgic and Sonja Grgic (2006), 'Independent Comparative Study of PCA, ICA, and LDA on the FERET Data Set', Wiley Periodicals, Inc.

[18] Dr. H.B.Kekre, Sudeep D. Thepade and Akshay Maloo (2010), 'CBIR Feature Vector Dimension Reduction with Eigenvectors of Covariance Matrix using Row, Column and Diagonal Mean Sequences', Int. J. of Computer Applications (0975 – 8887),Vol. 3, No.12.

[19] Sebastian Raschka (2014), 'Linear Discriminant Analysis bit by bit'.

[20] Shereena V.B. and Julie M. David (2014), 'Content Based Image Retrieval : Classification Using Neural Networks', The Int. J. of Multimedia & Its Applications (IJMA) Vol.6, No.5. DOI : 10.5121/ijma.2014.6503 31.

[21] Anca Apatean, Alexandrina Rogozan , Simina Emerich , Abdelaziz Bensrhair, 'Wavelets As Features For Objects Recognition'.

[22] Prof. C. S. Gode, Ms. A. N. Ganar(2014), 'Image Retrieval by Using Color, Texture and Shape Features', Int.Journal of Advanced Research in Electrical,Electronics and Instrumentation Engineering, Vol. 3, Issue 4, ISSN (Print) : 2320 – 3765,ISSN (Online): 2278 – 8875.

[23] Hamed Fadaei, and Thotsapon Sortrakul(2014), 'Content-Based Image Retrieval System with Combining Color features and Gradient feature', Int. Conference on Advanced Computational Technologies & Creative Media.

[24] G. Sasikala , R. Kowsalya and Dr. M. Punithavalli (2010), 'A Comparative Study Of Dimension Reduction Techniques For Content-Based Image Retrieval', The Int. J. of Multimedia & Its Applications, Vol.2, No.3, Doi : 10.5121/Ijma.2010.2303 40.

[25] R.Priya, Dr.Vasantha Kalyani David(2012), 'Enhanced Content Based Image Retrieval Using Multiple Feature Fusion Algorithms', Int. J. of Scientific & Engineering Research, Vol. 3, Issue 2, ISSN 2229-5518.

[26] K. Arthi, Mr. J. Vijayaraghavan (2013) 'Content Based Image Retrieval Algorithm Using Color Models', Int. J. of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 3, ISSN (Print) : 2319-5940 ISSN (Online) : 2278-1021.

AUTHORS

Shereena V.B.received her MCA degree from Bharathidasan University, Trichy, India in 2000. During 2000-2004, she was with Mahatma Gandhi University, Kottayam, India as Lecturer in the Department of Computer Applications. Currently she is working as Asst. Professor in the Department of Computer Applications with MES College, Aluva, Cochin, India. She has published papers in International Journals and International and National Conference Proceedings Her research interests include Data Mining, Machine Learning and Image Processing.

Dr. Julie M. David completed her Masters Degree in Computer Applications and Masters of Philosophy in Computer Science in the years 2000 and 2009 in Bharathiyar University, Coimbatore, India and in Vinayaka Missions University, Salem, India respectively. She has also completed her Doctorate in the research area of Artificial Intelligence from Cochin University of Science and Technology, Cochin, India in 2013. During 2000-2007, she was with Mahatma Gandhi University, Kottayam, India, as Lecturer in the Department of Computer Applications. Currently she is working as an Assistant Professor in the Department of Computer Applications with MES College, Aluva, Cochin, India. She has published several papers in International Journals and International and National Conference Proceedings. Her research interests include Artificial Intelligence, Data Mining, and Machine Learning. She is a life member of International Association of Engineers, IAENG Societies of Artificial Intelligence & Data Mining, Computer Society of India, etc. and a Reviewer of Elsevier International Journal of Knowledge Based Systems. Also, she is an Editorial Board Member of various other International Journals. She has coordinated various International and National Conferences.