# Face Recognition System Using: LDA and GMM based Approach

Aditi Mandloi
Shri Vaishnav Institute of Technology and Science
Indore, India

Priyanka Gupta
Shri Vaishnav Institute of Technology and Science
Indore, India

## ABSTRACT
A wide variety of systems require reliable personal recognition schemes to either confirm or determine the identity of an individual requesting their services. The purpose of such schemes is to ensure that the rendered services are accessed only by a legitimate user, and not anyone else. It is widely acknowledged that the face recognition has played an important role in surveillance system as it doesn't need the object's cooperation. The actual advantages of face based identification over other biometrics are uniqueness and acceptance. As human face is a dynamic object having a high degree of variability in its appearance, that makes face detection a difficult problem in computer vision. In this work we presented a novel Face Recognition feature Extraction Mode based on the combination of Linear Discriminant Analysis (LDA) and Gaussian Mixture Model (GMM). The proposed LDA and GMM based feature Extraction Model is utilized to search the feature space for the optimal feature subset where features are carefully selected according to a well-defined discrimination criterion. For the betterment of the feature classification a KNN classifier is used. The classifier performance and the length of the selected feature vector are considered for performance evaluation using MATLAB in ORL face dataset.

## Keywords
Face Recognition, Feature selection, LDA, ORL Dataset, Authentication, Gaussian Mixture Model;

## 1. INTRODUCTION
The human face plays an important role in our social interaction, conveying people's identity. Using the human face as a key to security, biometric face recognition technology has received significant attention in the past several years due to its potential for a wide variety of applications in both law enforcement and non-law enforcement. A wide variety of systems require reliable personal recognition schemes to either confirm or determine the identity of an individual requesting their services. The purpose of such schemes is to ensure that the rendered services are accessed only by a legitimate user and not anyone else [1].

## 1.1 Face Recognition
It is the general opinion that advances in computer vision research will provide useful insights to neuroscientists and psychologists into how the human brain works, and vice versa. Face recognition is one of the most relevant applications of image analysis. It's a true challenge to build an automated system which equals human ability to recognize faces. Although humans are quite good identifying known faces, we are not very skilled when we must deal with a large amount of unknown faces. The computers, with an almost limitless memory and computational speed, should overcome human's limitations. The face recognition techniques, mainly work in three steps [2].
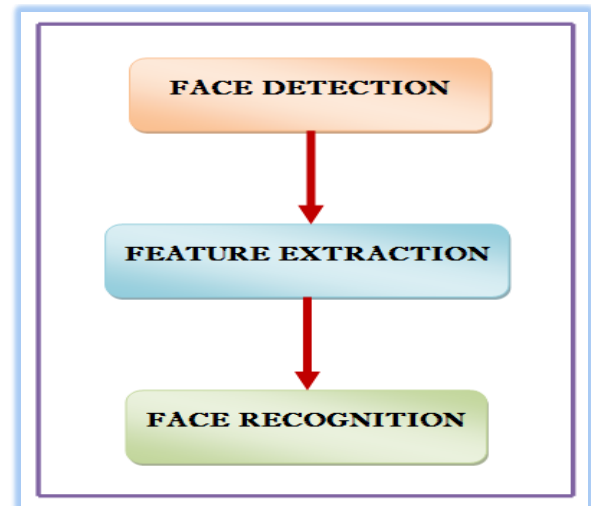


**Figure 1: Face detection and Recognition Overview**

    a.    Face Detection,

    b.    Feature Extraction,

    c.    Face Recognition.

a.    Face Detection

Face detection can be regarded as a specific case of object-class detection. In object-class detection, the task is to find the locations and sizes of all objects in an image that belong to a given class. Face detection can be regarded as a more general case of face localization. In face localization, the task is to find the locations and sizes of a known number of faces (usually one). In face detection, one does not have this additional information.

b.    Feature Extraction

The In pattern recognition and in image processing, feature extraction is a special form of dimensionality reduction. When the input data to an algorithm is too large to be processed and it is suspected to be notoriously redundant then the input data will be transformed into a reduced representation set of features. After the FD step, human-face patches are extracted from images. Directly using these patches for FR have some disadvantages, first, each patch usually contains over 1000 pixels, which are too large to build a robust recognition system. Second, face patches may be taken from different camera alignments, with different face expressions, illuminations, and may suffer from occlusion and clutter. To overcome these drawbacks, feature extractions are performed to do information packing, dimension reduction, salience extraction, and noise cleaning. After this step, a face patch is usually transformed into a vector with fixed dimension or a set of fiduciary points and their corresponding locations.

Transforming the input data into the set of features is called feature extraction [3].

c. Face Recognition

A facial recognition system is a computer application for automatically identifying or verifying a person from a digital image or a video frame from a video source. One of the ways to do this is by comparing selected facial features from the image and a facial database. It is typically used in security systems and can be compared to other biometrics such as fingerprint or eye iris recognition systems. Among the different biometric techniques, facial recognition may not be the most reliable and efficient. However, one key advantage is that it does not require the aid (or consent) from the test subject. Properly designed systems installed in airports, multiplexes, and other public places can identify individuals among the crowd.

*A. LDA*

LDA is widely used to find linear combinations of features while preserving class separability. Unlike PCA, LDA tries to model the differences between classes. Classic LDA is designed to take into account only two classes. Specifically, it requires data points for different classes to be far from each other, while points from the same class are close Consequently, LDA obtains difference projection vectors for each class. Multi-class LDA algorithms which can manage more than two classes are more used. Linear Discriminant Analysis (LDA) is most commonly used as a dimensionality reduction technique in the preprocessing step for pattern-classification and machine learning applications. The goal is to project a data set onto a lower-dimensional space with good class-separability in order avoids over fitting ("curse of dimensionality") and also reduce computational costs [4].

*B. Gaussian Mixture Model*

A Gaussian mixture model (GMM) is useful for modeling data that comes from one of several groups: the groups might be different from each other, but data points within the same group can be well-modeled by a Gaussian distribution. GMM provides a mathematical approach to statistical modeling of a wide variety of random phenomena and also provides a semi parametric framework to model an unknown probability density distribution shape. The combination of GMM with linear projection schemes is becoming popular as it provides more than one transformation matrix. Gaussian Mixture Models can be used to represent probability density functions, complex, from the marginalization of joint distribution between observed variables and hidden variables. Gaussian mixture model is based on the fact that a significant number of probability distributions can be approximated by a weighted sum of Gaussian functions [5] [6].

A Gaussian mixture model is a weighted sum of k component Gaussian densities for vector x as given by the equation,

$$p(x) = \sum_{j=1}^{k} \alpha_j \, p(x/j)$$

Where x is a D-dimensional continuous-valued data vector (i.e. Measurement or features)

$\alpha_j = 1,2,\dots k$, are the Mixture Weights

$p(AA\,x/j)$ = Gaussian Density Model for $j^{th}$ component,

Each component density is a D-variate Gaussian function of the form,

$$p(x/j) = \frac{1}{(2\pi)^{1/d}} \, e^{-1/2(x-\mu_j)^T \Sigma j^{-1}(x-\mu_j)}$$

With mean vector $\mu_j$ and covariance matrix $\Sigma j$. The mixture weights satisfy the constraint that $\sum_{i=1}^{k} \alpha_j = 1$.

Model parameter estimation is performed using a well known iterative method called Expectation Maximization (EM) which assumes that the number of components is known beforehand.

The complete Gaussian mixture model is parameterized by the mean vectors, covariance matrices and mixture weights from all component densities. These parameters are collectively represented by the notation $\mu_j$ and $\Sigma j$

## 2. LITERATURE SURVEY

Through the past three decades the face recognition systems have been developed. Indeed, this development depending on three kinds of recognition algorithms, namely profile, frontal, and view tolerant recognition, based on the type of images and the recognition algorithms.

The Gaussian mixture model (GMM) super vector approach is a well known technique in the domain of speech processing, e.g. speaker verification and audio segmentation. In this paper *Martin Schulz et al. [7].* Apply this approach to video data in order to recognize human facial expressions. Three different image feature types (optical flow histograms, orientation histograms and principal components) from four pre-selected regions of the human face image were extracted and GMM supervectors of the feature channels per sequence were constructed. Support vector machines (SVM) were trained using these supervectors for every channel separately and its results were combined using classifier fusion techniques. Thus, the performance of the classifier could be improved compared to the best individual classifier.

*Wen Wang et al. [8]* Presents a method named Discrimin1ant Analysis on Riemannian manifold of Gaussian distributions (DARG) to solve the problem of face recognition with image sets. Our goal is to capture the underlying data distribution in each set and thus facilitate more robust classification. To this end, we represent image set as a Gaussian Mixture Model (GMM) comprising a number of Gaussian components with prior probabilities and seek to discriminate Gaussian components from different classes. In the light of information geometry, the Gaussians lie on a specific Riemannian manifold. To encode such Riemannian geometry properly, we investigate several distances between Gaussian and further derive a series of provably positive definite probabilistic kernels. Through these kernels, a weighted Kernel Discriminant Analysis is finally devised which treats the Gaussians in GMMs as samples and their prior probabilities as sample weights. The proposed method is evaluated by face identification and verification tasks on four most challenging and largest databases, YouTube Celebrities, COX, YouTube Face DB and Point-and-Shoot Challenge, to demonstrate its superiority over the state-of-the-art.

In this paper, *Najim DEHAK et al. [9]* present a new face recognition algorithm. It supposes that a video sequence of a person is available both at enrollment and test time. During enrollment, a client Gaussian Mixture Model (GMM) is adapted from a world GMM using eigenface features extracted from each frame of the video. Then, a Support Vector Machine (SVM) is used to find a decisio11n border between the client GMM and pseudo impostors GMMs. At

test time, a GMM is adapted from the test video and a decision is taken using the previously learned client SVM. This algorithm brings a 3.5% Equal Error Rate (EER) improvement over the Bio 41+6+33-+Secure reference system on the Pooled protocol of the BANCA database.

Low-dimensional feature representation with enhanced discriminatory power is of paramount importance to face recognition (FR) systems. Most of the traditional linear discriminant analysis (LDA) -based methods suffer from the disadvantage that their optimality criteria are not directly related to the classification ability of the obtained feature representation. Moreover, their classification accuracy is affected by the "small sample size" (SSS) problem which is often encountered in FR tasks. In this short paper, *Kostantinos N, et al. [10]* Propose a new algorithm that deals with both of the shortcomings in an efficient and cost effective manner. The proposed here method is compared, in terms of classification accuracy, to other commonly used FR methods on two face databases. Results indicate that the performance of the proposed method is overall superior to those of traditional FR approaches, such as the Eigenfaces, Fisherfaces, and D-LDA methods.

With the goal of matching unknown faces against a gallery of known people, the face identification task has been studied for several decades. There are very accurate techniques to perform face identification in controlled environments, particularly when large numbers of samples are available for each face. However, face identification under uncontrolled environments or with a lack of training data is still an unsolved problem. *William Robson Schwartz et al. [11]* employ a large and rich set of feature descriptors (with more than 70 000 descriptors) for face identification using partial least squares to perform multichannel feature weighting. Then, we extend the method to a tree-based discriminative structure to reduce the time required to evaluate probe samples. The method is evaluated on Facial Recognition Technology (FERET) and Face Recognition Grand Challenge (FRGC) data sets. Experiments show that our identification method outperforms current state-of-the-art results, particularly for identifying faces acquired across varying conditions.

## 3. PROPOSED WORK

The proposed system is designed in two different data sets in which one data set passes to process training to the LDA and GMM and the other is passed for testing on LDA and GMM. Here in the training module system accepts user data in form of Image and calculate features. In testing phase system accepts user ORL data set in the form of image and the system produces features. On the basis of these calculated features we classify features by using K-nearest neighbor algorithm. The basic overview of the system is given below.

## 3.1 Methodology

The working of the proposed algorithm is given and explained in details in this section. To understand the core concept of the proposed methodology by using given figure 2 descriptions:
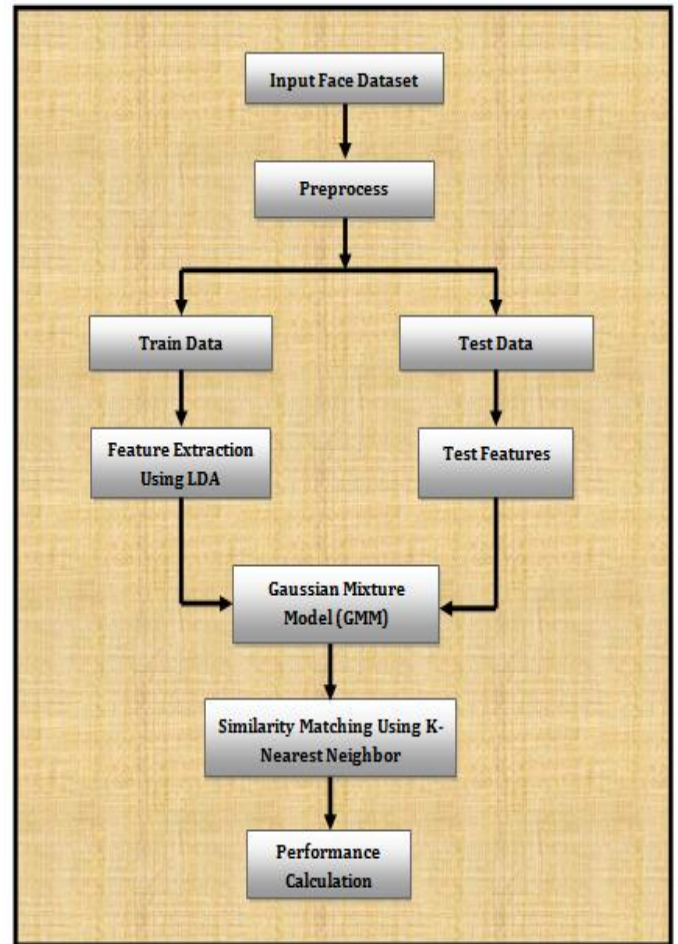


**Figure 2: Proposed Working Flow**

❖ **Input face dataset:** To process the entire project we need to load the face dataset. In this we give the input as ORL data set. While loading the input data set that process for the next step and initialize the process.

❖ **Preprocess:** This is done to enhance the images to improve the recognition performance of the system. After this process, our data set split into two different dataset, i.e. train data set and test data set. Different function performs on this dataset.

❖ **Train Data:** This is the data on which we performed training to train the model. On the basis of this model we process further steps

❖ **Test Data:** Similarly this is the data on which we performed testing process to test the model. For this testing is done measure for accuracy of the entire system

❖ **Feature Extraction using LDA:** This module finds the key features in the face that will be used in classification. It is responsible for composing a feature vector that is well enough to represent the faces. For this purpose we apply standard LDA algorithm for extracting the image features.

❖ **Test Feature:** This is a test feature which is calculated from the test dataset.

❖ **Gaussian Mixture Model:** On this model, we give input feature which is calculated by LDA along with give test features to process the function. GMM used for modeling the shape and texture of face images.

- ❖ **Similarity matching using KNN:** By modeling of GMM, we get optimal extracted features of this input image. Likewise, NN classify the different classes of similar attribute basis. NN classify the image faces of different 40 classes. That class contains 400 face images, i.e. each class contains 10 face images.

- ❖ **Performance Calculation:** Finally on the basis of the proposed working flow we can conclude that system using efficient performance of the face recognition basis of the following parameter 1) Recognition Rate, 2) Verification Rate, 3) Error rate.

# 4. PROPOSED ALGORITHM

The given section provides the detailed understanding about the proposed algorithm face recognition system:

**Table 1: Proposed Algorithms**

**Input:** ORL face Dataset,

**Output:** Performance Parameter for classification Accuracy

*Process:*

**1:** Load input face dataset (ORL)

**2:** Bifurcate data into train dataset and test dataset

**3:** Feature Extraction using LDA method. Following are the LDA steps:

- Suppose there are C classes

- Let $\mu_i$ be the mean vector for class $i$, $i = 1, 2, 3, 4, \ldots C$

- Let $M_i$ be the number of samples within class $i$, $i = 1, 2, 3, 4, \ldots C$

- Let $M = \sum_{i=0}^{C} M_i$ be the total number of samples, and

  Within Class Scatter Matrix : $S_w = \sum_{i=1}^{C} \sum_{j=1}^{M_i} (y_j - \mu i) y_j - \mu i T$

  Between Class Scatter Matrix: $S_b = \sum_{i=1}^{C} \sum_{j=1}^{M_i} (\mu_i - \mu)(\mu_i - \mu)^T$

  $\mu = \frac{1}{C} \sum_{i=1}^{C} \mu_i$ (Mean of entire dataset)

- LDA computes a transformation that maximizes the between-class scatter while minimizing the within-class scatter:

$$maximize \frac{\det(S_b)}{\det(S_w)}$$

- Such a transformation should retain class separability while reducing the variation due to sources other than identity (e.g., illumination).

**4:** Calculate test feature of the test data set

**5:** Apply GMM on training and testing features. Following are the GMM steps:

  Probability density function (PDF) with $k$ component for feature vector $x$ is

- $p(x) = \sum_{j=1}^{k} \alpha_j \, p(x/j)$ // $\boldsymbol{\alpha_j}$ = *Mixture Weight,*

  // $\boldsymbol{p(x/j)}$ =*Gaussian Density Model for $\boldsymbol{j^{th}}$ Component*

- $p(x/j) = \frac{1}{(2\pi)^{1/d}} e^{-1/2(x-\mu_j)^T \Sigma j^{-1}(x-\mu_j)}$ //$\boldsymbol{\mu_j}$= *Mean*

  *Vector,*

  //$\boldsymbol{\sum j}$= *Covariance Matrix for $j^{th}$ component*

- ☐ $\mu_j$ and $\sum j$ have to be calculated using Expectation Minimization Function

- ☐ $p(j/x_i) = \frac{\alpha_j \, p(x_i/j)}{\sum_{i=1}^{k} \alpha_i \, p(x_i/t)}$

- ☐ $\alpha_j = \frac{\sum_{i=1}^{n} p(j/x_i)}{n}$

- ☐ $\mu_j = \frac{\sum_{i=1}^{n} p(j/x_i) x_i}{\sum_{i=1}^{n} p(j/x_i)}$

- ☐ $\sum j = \frac{\sum_{i=1}^{n} p(j/x_i)(x_i-\mu_j)((x_i-\mu_j)^T}{\sum_{i=1}^{n} p(j/x_i)}$

**6:** Perform Similarity matching using K-nearest neighbor

**7:** Compute System Performance

**8:** Return Classification Accuracy

**9:** *End Process*

# 5. CONCLUSION

Face recognition is a very interesting quandary. Ideally a face detection system should be able to take a new face and return a name identifying that person. Mathematically, what possible approach would be robust and fairly computationally economical? If we have a database of people, every face has special features that define that person Face recognition is a technology just reaching sufficient maturity for it to experience a rapid growth in its practical applications. Optimal feature selection (FS) is to find the minimal subset of features from original feature set that can represent the whole dataset.

# 6. REFERENCES

[1] Rein-Lien Hsu, "Face Detection and Modeling for Recognition," PhD thesis, Department of Computer Science & Engineering, Michigan State University, USA, 2002

[2] B. Maison,C. Neti, and A. Senior, Audio-visual Speaker Recognition for Video Broadcast News: Some Fusion Techniques, In Multi-media Signal Processing, 1999.

[3] Wilson, Phillip Ian, and John Fernandez, "Facial feature detection using Haar classifiers", Journal of Computing Sciences in Colleges 21.4 (2006): 127-133.

[4] Sebastian Raschka, "Linear Discriminant Analysis", available online at: http://sebastianraschka.com/Articles/2014_python_lda.html#summarizing-the-lda-approach-in-5-steps

[5] Mitra, Sinjini. "Gaussian mixture models for human face recognition under illumination variations." (2012).

[6] Hassanpour, Reza, Asadollah Shahbahrami, and Stephan Wong, "Adaptive Gaussian mixture model for skin color segmentation", World Academy of Science, Engineering and Technology 41 (2008): 1-6.

[7] Wang, Wen, et al, "Discriminant analysis on Riemannian manifold of Gaussian distributions for face recognition with image sets", Proceedings of the IEEE

Conference on Computer Vision and Pattern Recognition. 2015.

[8] Schels, Martin, and Friedhelm Schwenker, "A multiple classifier system approach for facial expressions in image sequences utilizing GMM supervectors", Pattern Recognition (ICPR), 2010 20th International Conference on, IEEE, 2010.

[9] Bredin, Hervé, Najim Dehak, and Gérard Chollet, "GMM-based SVM for face recognition", Pattern Recognition, ICPR 2006, 18th International Conference on, Vol. 3, IEEE, 2006.

[10] Lu, Juwei, Konstantinos N. Plataniotis, and Anastasios N. Venetsanopoulos, "Face recognition using LDA-based algorithms", IEEE transactions on neural networks 14.1 (2003): 195-200.

[11] Schwartz, William Robson, et al. "Face identification using large feature sets." IEEE Transactions on Image Processing 21.4 (2012): 2245-2255.