Face Recognition Using Hybrid Techniques

By

Hafiz Salman Ashraf

MSEE-23

A thesis submitted to the faculty of Electrical Engineering Department, Military College of Signals, National University of Science and Technology, Islamabad, Pakistan in partially fulfillment for the degree of MS in Electrical (Telecommunication) Engineering.

March 2021

Dedication

This thesis is dedicated to my family and friends for their endless support, love and encouragement.

Acknowledgment

First of all I wish to express my deepest thanksgiving to my supervisor, Col (R) Dr. Imran Touqir for his boundless help, excellent supervision and kind suggestions throughout my studies and research work. I regard that it would be a great chance and opportunity for a student to work with him. Besides my advisor, I would like to thank the rest of my thesis committee: Brig Adil Masood Siddiqui (PhD), Maj Ajlaan Bin Mamoon for their insightful comments and encouragement, but also for the hard question which incented me to widen my research from various perspectives.

I thank my fellow lab mates in for the stimulating discussions, for the sleepless nights we were working together before deadlines, and for all the fun we have had in the last two years. Also I thank my friends especially Danish Ali and Idrees Ajmal for all the efforts they made for me and all the boundless help throughout my thesis from them will always be appreciated. I would like to thank my friends especially Usman Khan and Muhammad Hannan Khan and my senior Khadija Azmat for accepting nothing less than excellence from me also their encouraging words and endless support always motivated me.

Last but not the least, I would like to thank my family: my parents and to my brother and sister for supporting me spiritually throughout writing this thesis and my life in general.

Abstract

In real life applications whether it is government use (Law enforcement, Security/Counterterrorism, Immigration or Voter verification) or commercial use (Residential security, banking using ATM or physical access of buildings areas, doors or cars) identification of individual subject has become necessary for authentication purpose. Most of the techniques have been used from authentication perspectives like password based authentication or pin code based authentication for a subject in banking devices and smart phones. Some drawbacks also appeared for using these techniques as there was a possibility of password or pin codes to be stolen or hacked so, from security point of view these techniques might not be as good as it would have to be. Another technique named Finger print authentication has been used by security agencies and in the telecom sector as well, where thumb mark of an individual is essential and also for biometric verification. Another technique for identification named as Face Recognition has shown its role in areas of research and from the implementation point of view. In this technique face of an unknown subject is being first detected and then recognized after classification and then by comparison of it with other faces being stored in database. A Face image is full of information but using all the information will be time consuming and less efficient, so there is a need to use some of features of faces for identification and recognition purposes. This method has found its implementation in law enforcement agencies. For pattern recognition perspective the feature extraction is the most important task where we are required to obtain the best matching features with low processing time and for the best results. Another perspective to have concern is to reduce the dimensionality and complexity in the process that may also lead to reduce the time to process. In research perspectives several techniques has been proposed by different people in this method in different era for better recognition results. In this research work some existing techniques are merged to get the recognition results by using the methods of Principal component analysis (PCA) and SVD along with Hidden Markov Model (HMM) also the mean square error and PSNR are calculated.

Table of Contents

1 Introduction 1

1.1 Motivation 1

1.2 Types of Biometrics 2

1.2.1 Visual Biometric 2

1.2.2 Chemical Biometric 3

1.2.3 Behavioral Biometric 3

1.3 Face Recognition 3

1.3.1 Feature Extraction 3

1.3.2 Classification 3

1.4 Objectives of Thesis 4

1.5 Contributions 5

1.6 Thesis Outline 5

2 Literature Review 6

2.1 Feature Extraction Techniques 7

2.1.1 Geometry Based Technique 7

2.1.2 Template Based Technique 8

2.1.3 Appearance Based Technique 9

2.2 Subspace Methods 9

2.2.1 Principal Component Analysis (PCA) 9

2.2.1.1 Advantages of PCA 12

2.2.2 Linear Discriminant Analysis (LDA) 13

2.2.2.1 Advantages of LDA 15

2.2.2.2 Disadvantages of LDA 15

2.2.3 Independent Component Analysis (ICA) 16

2.2.4 Singular Value Decomposition (SVD) 16

2.2.4.1 SVD Subspaces 18

2.2.4.2 SVD Architecture 18

2.2.4.3 PCA versus SVD 18

2.2.4.4 Properties of SVD 18

2.2.4.5 Applications of SVD 19

2.3 Statistical Analysis 19

2.3.1 Mean Square Error (MSE) 19

2.3.2 Peak Signal To Noise Ratio (PSNR) 20

3 Hidden Markov Model (HMM) 22

3.1 Markov Property 22

3.2 Problem Associated with HMM 25

3.2.1 Evaluation Problem 26

3.2.1.1 Forward Algorithm 26

3.2.2 Decoding Problem 28

3.2.2.1 Viterbi Algorithm 30

3.2.3 Learning Problem 31

3.3 Advantages of Hidden Markov Model 32

4 Methodology 33

4.1 Proposed Methodology 33

5 Results and Conclusion 37

5.1 Recognition Rates with 6 states 37

5.2 Recognition Rates with 7 states 38

5.3 Peak Signal to Noise Ratio (PSNR) 42

References **44**

List Of Figures

2.1	Geometric Feature Extraction Technique	8
2.2	Extracted Images	8
2.3	Training and Testing phases of PCA	12
3.1	Transition and emission probabilities of three states	24
3.2	Trellis Path of hidden states	28
4.1	Recommended System	34
4.2	Proposed System	35
4.3	Seven States of HMM and their Transition Probabilities	35
4.4	Seven states of HMM for facial image	36
5.1	Test images successfully/unsuccessfully recognized (a)	40
5.2	Test images successfully/unsuccessfully recognized (b)	41
5.3	PSNR Graph	42

List of Tables

2.1 Features of Principal Component Analysis (PCA) 13

3.1 Different colored ball example 23

3.2 Probabilities of transitions of picking a colored ball 23

3.3 Emission probabilities of visible symbols 28

3.4 Probabilities of hidden states 29

3.5 Most probable path sequence for visible sequence for visible symbols of hidden states 29

5.1 Recognition Rates for 6 states 38

5.2 Recognition Rates for 7 states 38

Acronym

PCA	Principal Component Analysis
SVD	Singular Value Decomposition
LDA	Linear Discriminant Analysis
ICA	Independent Component Analysis
HMM	Hidden Markov Model
a_{ij}	Transition Probability from one state (i) to another state (j) in HMM
b_{jk}	Emission Probability of visible symbols in HMM
$\alpha_i(t)$	Probability to stay at hidden state at time step t
ω	Hidden States of HMM
V	Visible Symbols in HMM
γ	Refined Value for a_{ij} and b_{jk}
MSE	Mean Square Error

PSNR	Peak Signal to Noise Ratio
λ	Eigen Value
μ	Mean Value
DCT	Discrete Cosine Transform
DWT	Discrete Wavelet Transform
DFT	Discrete Fourier Transform
ANN	Artificial Neural Networks
KLT	Karhunen-Loéve Transform

Chapter 1

Introduction

1.1 Motivation:

Authentication of a subject is the most important aspect from security perspectives as it has found its main applications in real life such as Law enforcement agencies. Human identification is necessary when there is access of people in a restricted area and limitation of crowd required so, identification is required. Face recognition is prominent technique as compared to some existing techniques because of its non-intrusive property this was selected for my research purpose. Face recognition has become a prominent task in the field of research and implementation over the past few years. Recognition algorithms found their applications in the cell phones, law enforcement, security/counterterrorism, and immigration or voter verification. The reason is that the identification of a subject is necessary from security point of view. Many existing work has been done in this area of research and found their effective results, so with the process of time some advance techniques has also been found to get the better and effective results. Some techniques like password/pin code authentication and also the finger print authentication has been there in this field of research but face recognition has been found to be the best for its simple and robust approach. Finger print identification has been used for a long time and has been prominent in the area of research and applications. Finger print identification requires involvement of subject is essential which is to be targeted.

Face image is full of information but using all the information will be time consuming and less efficient, so there is a need to use some of features of faces for identification and recognition purposes. While having the features of any subject there is need to take some care for choosing as only those features are to be considered which can give the advantage of better results.

To perform this task some of the properties are to be considered for better results and performance:

- Format of pictures like sizing, color (grey scale/colored) and type of files taken
- No. of subjects in the database selected
- No. of images of an individual in the database
- Lighting conditions variations
- Positions of individuals in images
- Features variations

1.2 Types of Biometrics

Biometric is the measurement as well as statistical analysis of a subject's physical and behavior characteristics. It leads towards the study of biological statistical measurements of a subject. In some areas biometric relates to the authentication of a person in a way that who is he/she or what he/she is declared to be in this manner.

Some categories of biometrics are as follows:

1.2.1 Visual Biometric: It contains the categorization and measurements of faces of a subjects, which will lead further towards face recognition, Fingerprint Recognition, Signature Recognition, Eyes-Iris Recognition and Eyes - Retina Recognition, Hand Geometry Recognition and also Typing Recognition etc.

1.2.2 Chemical Biometric:

1.2.2 Chemical Biometric: In this type of biometric measurements of chemical cues are done such as Odor and DNA.

1.2.3 Behavioral Biometric: Performance of people in their daily life is measured in this category which contains speech analysis and gait analysis etc.

1.3 Face Recognition

Face recognition is prominent technique when compared with some existing techniques because of its non-intrusive property this was selected for my research purpose. So the complete process includes the detection and then the recognition while classifying it after using some suitable classifier where the processing time and the recognition rate have to be considered. Process starts with the image to be processed, to be detected and then to be recognized using some algorithm. Image that is to be processed is said to be a test image and where it is classified is said to have training images and that is known to be a data set as a combine form. Images those are stored in databases are knows as *training images*. Process in which certain algorithms are applied to train the images is knows as *training phase procedure*. Similarly some algorithms applied to create test images is said to be *testing phase procedure*. There may also exist some of the problems during this process like the variation of pose in an image, occultation of any feature in an image, conditions in an image and some other like resizing and quality of an image and also the illumination problems may be there and to fix these problems some of the technique are used.

Process consists of two steps in which we are to extract some of the features from the images and then we are required to classify them for proper recognition.

1.3.1 Feature Extraction

Usually images of a subject have been seen larger in size and also there is need to preprocess these images before actual processing for some better recognition results. Most of the informative values are to be taken from original data and also some of the pixel values may not be involved in positively contribution for recognition. Some existing techniques for extraction of features are: Principal Component Analysis (PCA) [20], Discrete Cosine Transform (DCT) [21], Singular Value Decomposition (SVD), Discrete Fourier Transform (DFT), Discrete Wavelet Transform (DWT) and Linear Discriminant Analysis (LDA). It has also been observed that by varying the illumination conditions the features extracted from SVD are not supposed to influence. Feature extraction starts from the block extraction which is then further divided into blocks of same sizes those are overlapping and then face recognition is applied on each of these blocks.

1.3.2 Classification

After extraction of features there is need to classify the images for recognition and most of the techniques have been applied for this objective like Artificial Neural Network (ANN), Support Vector Machine (SVM), Euclidean Distance and also Hidden Markov Model etc. Hidden Markov Model has been used in most of the area of research and application like in word, speech and gesture recognition. HMM was applied in face recognition by Samaria and Fall side. HMM has been applied for different states at different times. Human face comes in order from top to bottom starts from Hair, forehead, eyebrow, eyes, nose, lips and chin, these regions of facial images can be modeled and applied in face recognition classification by using HMM.

1.4 Objectives of Thesis:

The main objectives of the thesis are:-

- Maximize the recognition rate and to reduce the processing time.
- Use a technique that can overcome the computational complexity and the dimensionality of larger data.
- Some suitable classifier for the classification purpose to produce better results.

1.5 Contributions:

Hybrid techniques are applied in this research work where dimensionality reduction and feature extraction were objectives and for this reason two different techniques were applied. Classification responsibility was given to HMM. In this research work efficient recognition results were obtained and also the graph of PSNR was also found. This combination of methods worked satisfactory results as the recognition results shows its efficiency.

1.6 Thesis Outline:

Thesis is composed of six chapters where;

- **Chapter 1** has introduction part where introduction to different authentication methods has been discussed and also the face recognition
- **Chapter 2** is nominated for literature review where some past methods in this field has been compiled and their results and efficiency was discussed.
- **Chapter 3** is presented for Hidden Markov Model (HMM) and its aspects and some problems associated to this classifier.
- **Chapter 4** is set for methodology that has been used in this task and also the technique that is being proposed here.
- **Chapter 5** is for results and conclusion where obtained results are displayed.
- **Chapter 6** is for the future work where some ideas are shared to work on this topic in future.

Chapter 2

Literature Review

Initially it was about Geometrical Matching methods where geometric features create the whole face. PCA approach applied to tackle the problem of face recognition problem was studied based on the Eigen faces approach which was proposed [1]. Facial expressions were recognized by using PCA and with SVD and results were reported for expression of happy 95%, expression of disgust 70%, expression of surprise 85%, expression of angry 60% and expression of sad 90%.

Distance measuring is also the most important part in face recognition and also different classifiers were tested with PCA algorithms, PCA algorithm with mean clustering have produced best results of recognition when applied with Squared Euclidean Distance as compared to the Euclidean Distance and also the city block method [4].

Hidden Markov model (HMM) is prominent classifier and has found its applications in areas of speech, gesture, word recognitions. HMM was used for face recognition by Samaria in 90s [9]. Later on different techniques like DCT and SVD were applied with HMM and computational complexity was reduced and rate was preserved when compared with Euclidean distance method [5, 6, 7].

HMM was applied with DWT, in which coefficients of DWT have been used for feature extraction and used as feature vector and then HMM applied there for classification purpose [8]. Facial expression has some important information for the states of expressions and facial regions [11]. Input images used for purpose of recognizing the expression of facial information need to identify the position of faces in image and also recognizing the expression [11]. Face recognition got the advantage for being universal over other major biometric features because a face image can be easily captured without the person’s knowledge such as using a surveillance camera, whereas fingerprints or some other biometric features like iris image, and palm prints are captured with much more difficulty and that cannot be captured with a good quality without the persons co-operation [10].

Most challenging task in this work to get the features those are likely to be obtained using some feature extraction techniques before it is further processed. Face detection is all about whether an image depicts the face that is the general structure of image which includes eyes, forehead, lips, eyebrows and nose etc. Face detection is easy task when compared with face recognition as in face recognition it is required to find whose face is this. Some of the variations may cause difficulties for recognition process like illumination change, pose variations, cluttering, occlusion and RST variation. Some techniques are used to recognize a testing image from its training images placed in dataset. These techniques are applied according to their effectiveness in different scenarios and different conditions. Techniques were applied in different scenarios and the recognition of images was obtained.

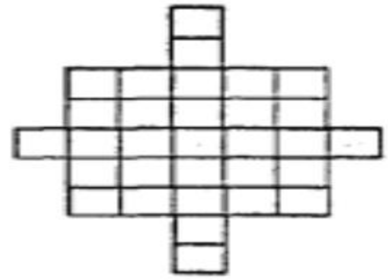
2.1 Feature Extraction Techniques

Some techniques being used in face recognition process are:

2.1.1 Geometry Based Technique

Techniques fall in this category are based when a set of geometrical features are computed [10]. The total geometrical configuration is described by a vector finding the position along with the size of facial features, like features of eyes, nose, mouth and eyebrows [10]. Each individual has different feature like the orientation and the location of nose, eyes, ears and lips. Features are extracted by using the size and also with the important components of relative position of images [10]. Using the minimum distance classifier the images are classified. In this technique, the full face image contour is taken and then subtraction from an unknown image is done. This is done by using the extraction window. Feature vector is being prepared after finding the edges and the direction of some important components [10].

(a) Face Image Contour



(b) Extraction Window

Fig. 2.1: Geometric Feature Extraction Technique

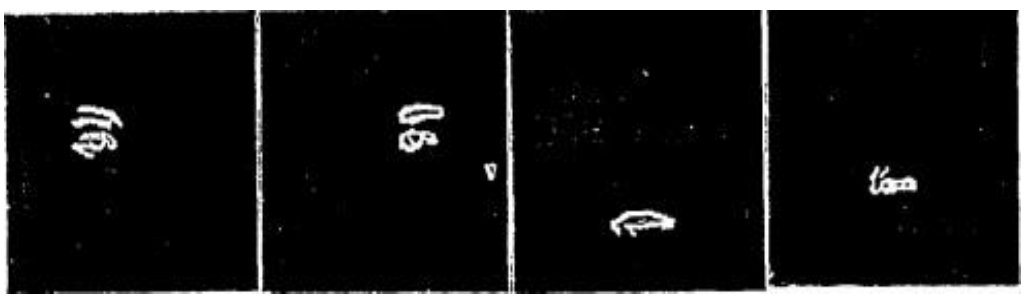


Fig. 2.2: Extracted Images

2.1.2 Template Based Matching

In template matching based techniques the test image can be represented by intensity values of two-dimensional array after it is compared by using a suitable metric for example Euclidean distance, where only single template of face image representing the whole face [10]. Template based techniques include detection of features of eye with mouth; initially template of an eye is being used to detect the area covered by an eye from image [10].

2.1.3 Appearance Based Technique

This method has the concept of features for all different from some simple facial features like features of eyes and nose. Some of characteristics are being extracted from the images to be considered as feature. This approach has been found to be one of best among others as it collects the most useful information and rejects the redundant information. Methods fall in this approach are Principal Component Analysis (PCA), Singular Value Decomposition, Independent Component Analysis (ICA) and Discrete Cosine Transform (DCT) are to be used to find the features from set of images. Main idea behind this is to overcome the large dimensionality to smaller dimensionality of observed variables for independent variables without the loss of important information. ICA is the technique which uses the information of independent component analysis for which they use the second-order statistics and high order statistics as well [10]. In DCT the set of coefficients is very large and hence it becomes complex to analyze and store such data [10]. Larger data can be difficult to process so, we need to reduce its dimensionality in this regard so as to simplify the calculations and to get the best results without complex calculations. Some techniques are used for extraction of features as well as to reduce the dimensionality. Suppose a picture having large size is very difficult to process and if we take some of its feature which can be used as whole for features can be a remedy to this problem.

2.3 Subspace Methods

In the process of face recognition, subspace methods leads towards the process that evaluates the basis vector of a project that optimally cluster the data according to their class or in other way subspace methods contains the properties of larger space which can be used in face recognition process. Some subspace methods are as follows:

2.3.1 Principal Component Analysis (PCA)

Principal component analysis (Karhunen-Loéve transform or hotelling transform) is one of most broadly used techniques in subspace methods. It can be known as the signal representation technique in which the facial basis are found which are also known to be the Eigen faces, these Eigen faces are obtained from covariance matrix which will lead towards the training set of images. PCA is an efficient statistical technique for testing the data presented in the classes. One of the basic purposes for the PCA is assumed the reduction of the dimensions for the observed data that has found a large number for the correlated variables without the loss of some important information. Distance measuring is also the most important part in face recognition and also different classifiers were tested with PCA algorithms, PCA algorithm with mean clustering have produced best results of recognition when applied with Squared Euclidean Distance as compared to the Euclidean Distance and also the city block method [4]. Gosavi and Khot applied PCA with SVD to get the expression recognition with 67% [17].

Mathematical form of PCA has three basic steps:

Initially transformation matrix is created; training images are aligned into the matrix vector in second step. Test image can be recognized by putting it into the subspace and then compared it with trained images placed in the dataset. Later on with this work in 1991 Turk and Petland progressed in this field [14, 15]. Eigen faces which can be found to overcome the dimensions are feature vector space leads towards Eigen vector, Eigen vectors are associated with the most dominant Eigen values from covariance matrix. After this most dominant component is assumed to be the first principal component and that one which has less information about the images is the second principal component and denoted as PC1 and PC2 respectively.

PCA can be seen in the following aspects:

a) Suppose each face image is represented by X(x,y), two dimensional (a × b) is explicated as one dimension (ab), also set of training facial images are $\{X_1, X_2, X_3 \ldots X_N\}$, by having average for training facial images we use,

$$\overline{X} = \frac{1}{N}\sum_{i=1}^{N} X_i$$

b) To calculate the Covariance Matrix:

$$C = \frac{1}{N}\sum_{i=1}^{N}(X_i - \overline{X})(X_i - \overline{X})^T$$

Mean subtracted values are considered and the covariance matrix is found from these values.

c) Find the eigen vector in accordance to eigen values by using:

$$CV = \lambda V$$

Most dominant term of Eigen value is assumed and according to that value the Eigen vector is calculated.

d) Mean values of each image centering to project into eigen space using:

$$W_i = V_i^T (X_i - \overline{X})$$

e) Testing of face images can be then compared with the data base of images in Eigen domain.

f) To calculate similarity measure, use the classifiers.

Karhunen-Love Transform faces two limitations in its process, Firstly, large size data may lead to reduce the feature discrimination that is in other words Eigen faces. Secondly, Computational complexities while finding the eigenvectors as well as storage requirements are enhanced with increasing training databases. Feature vector of an individual’s face image that is having the weights of images are calculated which will further help in finding the recognition process and recognition rate as well. Complete process using PCA and the minimum distance classifier is shown in the figure which will be having the training images and the test images and the weights of those images which are participated in calculating the recognition of images using this technique.

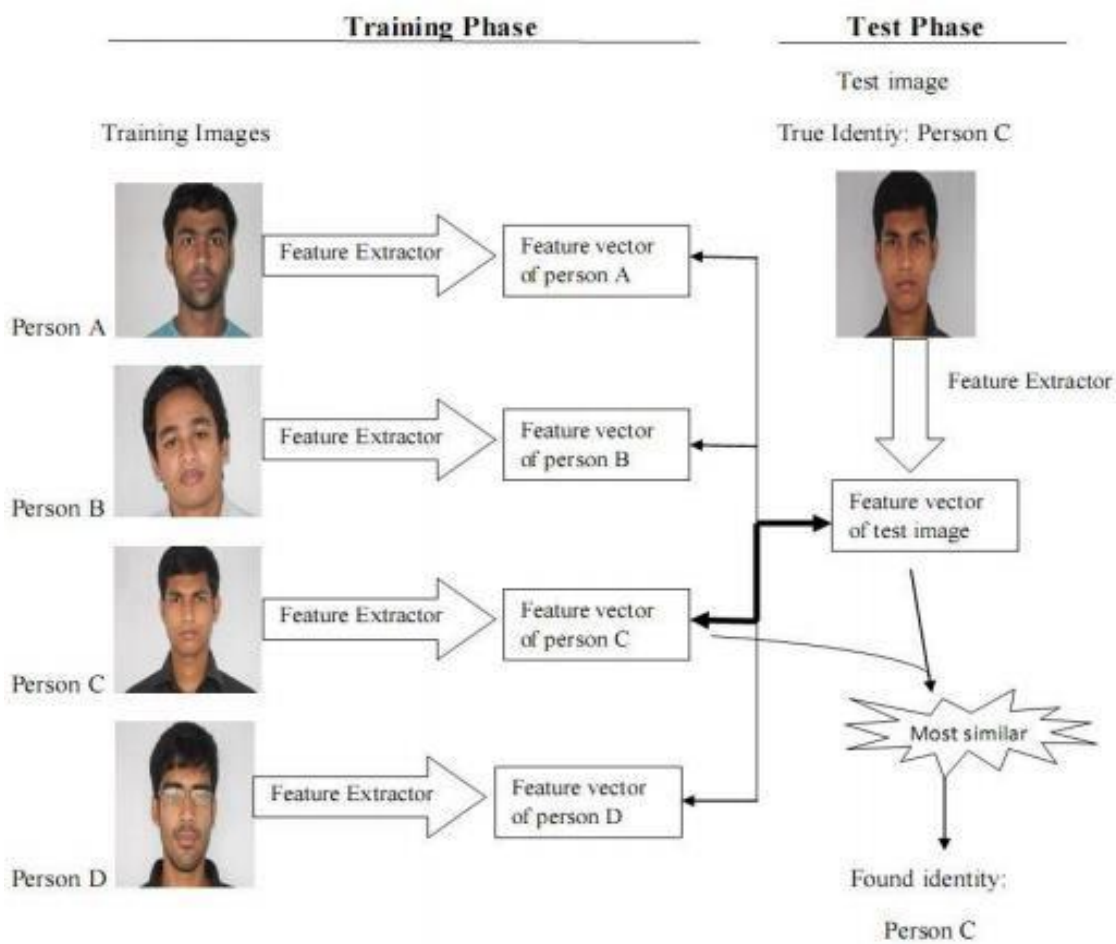


Fig. 2.3: Training and Testing phases of PCA

2.3.1.1 ADVANTAGES OF PCA:

- PCA has lack of redundancy for the data which is given in the orthogonal components.
- By using PCA complexity is reduced.
- Representation of database is smaller on a reduced basis.
- Noise can be reduced as the maximum variation basis is selected so the small variations are ignored automatically in the back ground.

The features of PCA are shown in the table 2.1 below:

Features	**Principal Component Analysis**
Distinction of classes	PCA aligns the whole data without taking it for examines basic class structure.
PCA Implementation	PCA is implemented in the prominent investigation fields of criminals are efficient
Direction of highest discrimination	Directions of highest distinction are different from directions having maximum variance
Focus of technique	PCA identifies the directions contains widest variations
Supervised learning method	PCA is unsupervised method.
Properly disperse classes for minor datasets	It is not as strong method as others.

2.3.2 Linear Discriminant Analysis (LDA)

LDA which is also known as *Fisher Faces* technique, is another subspace method in which supervised learning of data is done and the data is labeled as in the form of classes. Some of the basic difference between LDA and PCA is that in PCA data is obtained but can't be categorized or it can't be said that from which class the data is obtained. One of the main and basic differences between these two techniques is the labeling of data. Another difference is that in LDA it projects the data towards the maximum variation which is not done in PCA. In other words LDA always seeks the direction of maximum variation in dataset.

LDA is used to separate the data and its main purpose is to observe the direction that can properly differentiate the different data in classes. A human face is represented by data matrix X in which each row describes distinct human face. Every image is to be as X and (n,m) are to be the pixels and in high dimensional space as n*m. While we are using LDA we are interested for maximizing the separability between two known groups. Just like PCA, LDA can also be used to extract some features from images. The main differentiation between PCA and LDA is that, in PCA the extracted features don't represent the class where they belong to while in LDA classes are specified to describe the data from where features are obtained. Suppose if we have a graph of two classes in two dimensions there may be two ways of reducing it; a bad way and a good way. In bad way one of two dimensions are ignored and that's a bad thing because most of the important information is lost in this way and that may not be advantageous while processing the data, while the good way is that select a line in two dimensional data graph in which most of the information is found. Line which is drawn is said to be the new axis and all the values are projected upon it. If we optimize the distance between means and scatter, we can get some better results for separation. LDA based techniques can be used in pattern recognition, computer vision and also in the face recognition methods. It has found its dominant role in this field for research and implementation as well. Kapil performed with PCA and LDA for better recognition rate and processing time with complexity as well [16].

Main objectives in LDA are;

- Find the mean values of data sets
- Creation of an axis that can maximize the distance between means
- Minimize the variation (which LDA calls '*Scatter*') within each category

Both the criteria simultaneously represented as,

$$(\mu-\mu^*)^2 / s^2 + s^{*2}$$

Where μ and μ^* are the mean values of datasets, s and s^* represents the scatter of two datasets. s is squared and the main reason behind it is that it is unknown that which category/class is larger than the other and also we don't know the numbers to be negative. $(\mu-\mu^*)^2$ can also be represented by $\boldsymbol{d^2}$. Similarly if we have more than two dimensions, process is still the same. But if we have two axis and three categories then, find a point that is center to all data points and rest of the process is same. LDA supposed to be better than PCA but isn't true for small training data. In PCA and ICA the distance between weights of face of same subject is greater than between different subjects. Rectify this FISHERFACE method based on LDA was proposed in which vectors were found that can best discriminate between classes of data along describing data.

2.3.2.1 Advantages of LDA:

- Remedy to the above problem is to use incremental PCA for the purpose of updating the data even without completely retaining the images but there is still main difficulty is to identify the covariance matrix and then to tackle the inverse of within class scatter matrix for developing the incremental PCA/LDA
- 2D LDA/PCA can also be utilized to get covariance matrix without conversion of matrix to vector and that can be helpful for recognition purpose.
- LDA works better as the classes/categories are found in separation process.
- By using both PCA and LDA, first it may also be used to reduce dimensionality of large data set, then transformation is found to best for differentiating one class of images from the other

2.3.2.2 Disadvantages of LDA:

- Subspace methods are data dependent, for this reason collection of faces varies the basis vectors for projection should be recomputed which will lead to prominent workflow.
- One of challenges is that how dynamically incorporate fresh data and refine basis vector without retaining the whole data set after every new image acquisition.

2.3.3 Independent Component Analysis (ICA)

Independent Component Analysis (ICA) is also considered to be a subspace method and also as extended version of PCA. The main difference is that it only deals with the second order statistics as compared to PCA. ICA deals with higher order data in its calculation. ICA can be there for feature extraction techniques in face recognition like PCA, SVD and LDA. Basis images extracted in ICA are independent for training process as compared to PCA where images are uncorrelated. Support Vector Machine (SVM) and ICA both were introduced by Zhang.

2.3.4 Singular Value Decomposition (SVD)

SVD is one of feature extraction technique that can be used in face recognition to extract feature. SVD can be looked from three perspectives. From one perspective it can be process for transformation of correlated variables into the uncorrelated ones from where it can depict the versatile relationships with original data. From the second perspective it is method for finding and aligning the dimensions from where a data point shows the maximum variations. The third way of analyzing the SVD is when we have found where the most variation exists, it makes possibility to depict the best approximation using less dimensions of original data points. For most significantly SVD is a method of data reduction.

Just like PCA, SVD assumed to be given large data dimensions which can be reduced into lower dimensions and from where substructure of original can be seen more clearly by using less dimensions.

Theorem can be seen in the following equation:

$$A_{mxn}=U_{mxm}\, S_{mxn}\, V^{T}_{nxn}$$

Where $U^TU=I$ & $V^TV=I$

It can be seen that information and energy of a signal is mostly transferred by some of large singular values and their related vectors. To find some of coefficients as feature, large number of combinations of singular values and other participants of

U & V were extracted. It was found that SVD contains three matrixes (U, S and V) use of two first coefficients of matrix S and first coefficient of matrix U as three features U_{11} S_{22} S_{11} associating each block have the best classification rate [10].

Without using the large dimensional data SVD can give some better results by using only few of its dimensions which can depict the whole structure.

Let matrix A be an m x n. The matrix A^TA is to be symmetric and it can be diagonal zed. Evaluating the symmetric matrix of A^TA, two things are to make sure: The Eigen values of A^TA will be real and to be non-negative. The eigenvectors are to be orthogonal. To process two orthogonal matrices U and V that digitalizes an m x n matrix A [11]. To find the V, $A = U\Sigma V^T$ [11]. Eigen values from A^TA are decreasing ordered magnitude after rearranging [11].

$$[\overset{N}{X}]_M = {}_M[\overset{M}{U}]\ [\overset{N}{S}]_M\ [\overset{N}{V}]^T{}_N$$

- $U = [u_1, u_2 \ldots u_m]$ $\qquad\qquad$ $V = [v_1, v_2 \ldots v_n]$

$$S = \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & O & \\ & & & \sigma_n \end{bmatrix}$$

U matrix is an *M×M and* orthogonal, while *V matrix* is an *N×N and* orthogonal, and *S has M×N dimensions* and the diagonal elements in this represents the singular values, si of *X*. The superscript *T* is there for transpose of matrix. Columns in matrix *U* are left singular vectors, while columns of orthogonal matrix *V* are right singular vectors.

The left singular vector of *X* is eigenvectors of XX^T and the right singular vectors of *X* are eigenvectors of X^TX. U and V are unitary orthogonal matrices and S is a diagonal matrix of decreasing singular values [18].

In process of feature extraction SVD algorithm can be run for the computation of face features for the purpose of recognition, by calculation the U, V, S matrix. Facial expressions can be found and SV can be used as facial feature for face recognition. After this process these features are to send to the classifier for the classification of every given face query.

2.3.4.1 SVD Subspaces: SVD is composed from two orthogonal dominant and subdominant subspaces. This corresponds to separation of the M-dimensional vector space into *dominant* and *subdominant subspaces.* This smart property of SVD has found its application in noise filtering and watermarking [18].

2.3.4.2 SVD Architecture: Decomposition of an image in SVD, the singular values (SV) represents luminance of image for corresponding pair singular vectors reveals the geometry of the image layer. The largest object components in an image found using the SVD generally compatible to Eigen images associated with the largest singular values, while image *noise* compatible to Eigen images associated with the SVs [18].

2.3.4.3 PCA versus SVD: PCA is used to identify the dominant vectors showing a given data set and gives an optimal basis for minimum mean squared reconstruction [18]. Computational basis for PCA is the calculation of the SVD of the data matrix, or equivalently the Eigen values decomposition of the data covariance matrix SVD is closely corresponds to the standard Eigen values-eigenvector or spectral decomposition of a square matrix X, into VLV', where V is orthogonal, and L are diagonal. In fact U and V of SVD represent the eigenvectors for XX' and X'X respectively. If X is symmetric, the singular values of X are the absolute value of the Eigen values of X [18].

2.3.4.4 Properties of SVD:

- Rank Approximation
- Orthogonal Subspaces
- Image Denoising
- Image Compression
- Image Forensic

2.3.4.5 Applications of SVD:

In case of *Least Square Problems* SVD finds its one of application where in case we are given most of the points and we are to determine the curve that gives the best fitting solutions to those points. Similarly in case of Image Compression there is need to perform this action for proper transmission of pictures and data. Images are to be resized and compressed to enhance the efficiency of images in a screen to make it better fit. In some cases where we are given tons of data and the analysis and solution to that data is required then *SVD* plays an important role there to get a solution of it. Like *Face Recognition* there exists another technique named *Hand Written Recognition* and for its proper process we can implement SVD just like in *Face Recognition* by solving the images in the matrices forms. In some cases when the transmission of images takes place there are possibilities of data loss. So *SVD* can be implemented to for the reconstruction of images as we use this in *Hand Written Recognition*. Singular Value Decomposition is being used widely in image processing for feature extraction and computation of larger data in a suitable way.

2.4 Statistical Analysis

Process to gather, observe, summarize and analyze it by numerical formulas and calculate it is the statistical analysis. There are few most common and effective measures are for statistical analysis i.e. *Mean, Median*, *Variance* and *Standard Deviation (S.D)* etc. *Mean*, *Median* and *Mod* relates to the central point or tendency of a data set, these kinds of measures are related to some of errors in data sets. Error can be occurred during the collection, analysis and also for calculations of data and main purpose during the statistical analysis is to quantify the error. There are few methods to quantify the error in statistical analysis; *Mean Square Error* is one of them.

2.4.1 Mean Squared Error (MSE)

Concept of mean squared error is considered to be an important criterion for the analysis of performance of an estimator. Error is quantified by averaging the error of an estimator. Average squared difference between the values those are estimated and those which are actual values are calculated, this will lead towards the calculation of error during the whole process. MSE is considered as risk function. Mean squared error has the property of being a strictly positive value that will make it sure that the value will never be negative; the reason behind this property is that of randomness because an estimator can't account the information which can make sure a more accurate estimate.

Smaller the values of MSE the closer you are to the line of best fit. MSE intimates you that how close is the regression line to the data points. It is done by having a distance from regression line to the set of data points and further it averaged to eliminate the negative sign as the MSE always have positive values. Distances are the actual errors in the whole process.

There are few steps to calculate the MSE;

- Identify the regression line
- Put your values of X into equation (linear regression) and to identify the newly Y values
- Newly found Y values are subtracted from the original values, this will identify the error
- Squaring the error
- Adding up the error
- Calculate the mean

General form of MSE is;

$$MSE = \frac{1}{n}\sum_{i=1}^{n} (y_i - \bar{y}_i)^2$$

2.4.2 Peak Signal to Noise Ratio (PSNR)

Peak signal to noise ratio or PSNR is the ratio used for quality measurements between the two images (compressed and original). Also it is considered as the ratio between the signal's maximum possible power and the power for distorting noise of a signal. It is calculated in decibels. As the value of PSNR is higher, better is the result of compressed image.

MSE and PSNR both are used as quality matrices for image quality measurements. Higher the value of PSNR will lower the value of MSE and that of better quality of compressed images.

General form to calculate PSNR is;

$$\mathbf{PSNR = 10\ log_{10}\,(R^2 / MSE)}$$

R is assumed to be the highest variation in the input image

Chapter 3

Hidden Markov Model (HMM)

3.1 Markov Property

Hidden markov model is straightaway a statistical model that can generate model depending upon the input sequences. If the is given entry sequences this can give you transition probability. A **Markov chain** is "a stochastic model describing a sequence of possible events in which the probability of each event depends only on the state attained in the previous event e.g. Stock Market or Weather prediction etc. A Markov process advances from one state to other also depending only on the previous n states.

Human beings had been blessed with the ability to differentiate between faces of different people, so the computers are recently developed with this ability. Face recognition has found its popularity in various places like, protection, scrutiny, and biometric verification system, verification of laptop, mobile passwords as well as credit card and Image database management. One of the very famous and recent methods used for achieving Face recognition is the Hidden Markov Model (HMM). After detection process, test image is to be classified using one of different classifiers like Artificial Neural Network (ANN), Euclidean Distance Method or Hidden Markov Model (HMM). Stochastic process has Markov property that if the conditional distribution of future states depends on present value and it will also be independent of past values.

Suppose we have an example of choosing the color ball from an urn containing three different colors of balls, we chose the ball randomly and the probabilities associated with them are shown below:

Urn 1	**Urn 2**	**Urn 3**
# of Red Balls = 30	# of Red Balls = 10	# of Red Balls = 10
# of Green Balls = 50	# of Green Balls = 40	# of Green Balls = 40
# of Blue Balls = 20	# of Blue Balls = 50	# of Blue Balls = 50

Table 3.1 Colored ball example

From above example the probability of picking a ball and transition probability is shown in the table which depends upon given condition:

	U1	**U2**	**U3**
U1	*0.1*	*0.4*	*0.5*
U2	*0.6*	*0.2*	*0.2*
U3	*0.3*	*0.4*	*0.3*

Table 3.2 Probabilities of transition of picking a colored call

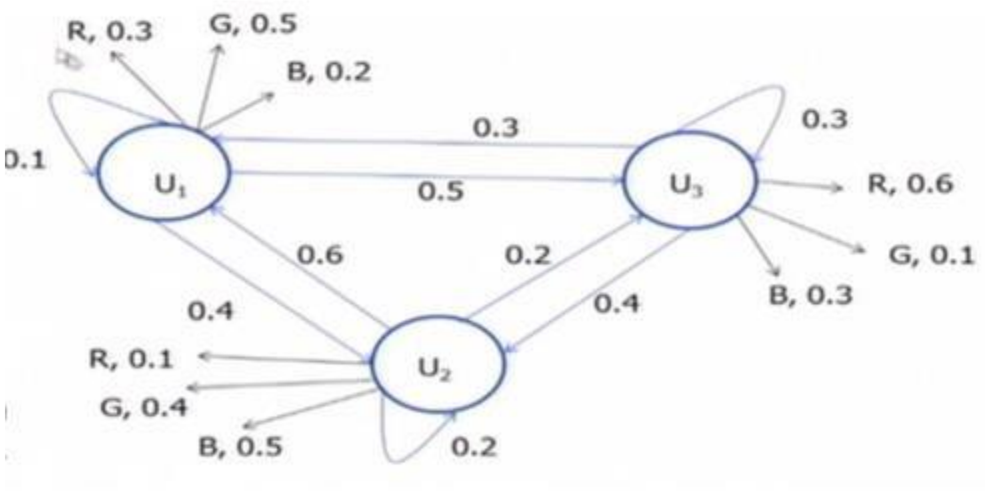


Fig. 3.1 Transition and Emission Probabilities of Three States

In 1994 Samaria and Young put their efforts to introduce HMM as a mechanism for recognition purposes. One of main features of HMM is that it segments the face image into different meaningful regions for better performance in recognition purposes. HMM is found to be a demographic system in which is to be modeled is assumed to be in hidden states.

HMM can’t observe a state directly so hidden states are considered to be connected.

HMM model is described as the following parameters:

- Set of states : S where $|S|=N$
- Output Alphabet : V
- Transition Probabilities : $A = \{a_{ij}\}$
- Emission Probabilities : $B = \{b_j(o_k)\}$
- Initial State Probabilities : π

$$\lambda = (A, B, \pi)$$

Hidden Markov Model (HMM) containing two sets of states and three sets of probabilities;

- **Hidden States:** A TRUE state of a system that narrates by a Markov process e.g., the weather.
- **Observable States:** These states are `visible' e.g., seaweed dampness.

3.2 Problems Associated with HMM:

Following are the problems related to HMM;

- Calculating the probability for a visible sequence given HMM *(Evaluation)*.
- Identifying hidden states sequence for which is most probable created by observed sequence and named as *Decoding*.
- Producing an HMM for sequence of observations (*Learning).*

Suppose we have hidden states $\omega = \{\omega_1\ \omega_2\ \omega_3\}$ and visible states $v = \{v_1\ v_2\ v_3\}$ of model $\boldsymbol{\theta}$.

Machine will make a transition from one state to the other and the transition probability is '$\mathbf{a_{ij}}$' that is denoted by;

$$\omega_i\ (t-1) \rightarrow \omega_j\ (t)$$

From ω_1 to ω_2 the machine will make a transition with probability a_{12}. Similarly from ω_2 to ω_1 machine will make a transition with probability a_{21} and so on. From state ω_1 the machine will emit a visible symbol v_1 with the emission probability b_{11}. For this statement we can write that a machine will emit v_k when it is at state ω_j with probability b_{jk}

$$P\ (v_k / \omega_j) = b_{jk}$$

From state ω_1 it will emit v_2 and v_3 with probabilities b_{12} and b_{13} respectively and same is the case with all other hidden states. Now finding given any state there will always be a transition to some of the states including the state itself. ω_1 will have a self-transition with probability a_{11} and same is the case with other states. Machines will always a make a transition so the probability of transition will be;

$$\sum_j a_{ij} = 1, \forall\ i$$

Machine will always a emit a visible symbol so the probability of emission will be;

$$\sum_k b_{jk} = 1, \forall\ i$$

HMM will always have a final state and whenever the machine will reach at that stage it cannot come back to any stage and it will emit only one visible symbol.

3.2.1 Evaluation Problem:

Consider a scenario where you face a number of HMMs those are showing different systems and also with the observation sequence. We are then interested to find which of the system has produced this sequence. Suppose we have two different models named as ‘summer’ and the ‘winter’, and both are for seaweed. As the behavior can be different from one season to the other, we can then expect to identify the season on the basis of the dampness observations. A forward algorithm can be present to find the probability of the observation sequence associated with HMM. So these kinds of problems occur during speech recognition where large number of HMMs is modeled for a particular word.

3.2.1.1 Forward Algorithm:

Suppose a HMM has the following parameters

$$\theta \rightarrow \omega, v, a_{ij,}\ b_{jk}$$

Visible symbols sequence called as v^T. Evaluation problem says given v^T and θ, $P(v^T/\theta)$ is to find

$$P(v^T/\theta) = \sum_{r=1}^{rmax} P(v^T)/(\omega_r^T).P(\omega_r^T)$$

$$\omega^T_r = \{\omega(1), \omega(2), \omega(3), \ldots \omega(T)\}$$

If there are N → no. of hidden states, $r_{max} = N^T$

$$P(\omega_r^T) = \prod_{t=1}^{T} P\,(\omega(t)/\,\omega(t-1)$$

The above equation can calculate a particular set of sequences for hidden states.

$$P(v^T/\omega_r^T) = \prod_{t=1}^{T} P\,(v(t)/\,\omega(t)$$

And also

$$P(v^T/\theta) = \sum_{r=1}^{rmax} \prod_{t=1}^{T} P\,(v(t)/\,\omega(t)\,P\,(\omega(t)/\,\omega(t-1)$$

Complexity of this solution is $N^T T$, so a high number of computations are required. So instead of using this complex solution we use a recursive algorithm which says that Hidden Markov Model is at hidden state at time step t, and it will produce visible symbols sequence.

$$a_j\,(t) = \begin{cases} 0, & t = 0 \text{ and } j \neq \text{initial state} \\ 1, & t = 1 \text{ and } j = \text{initial state} \\ \left[\sum \alpha_i\,(t-1)a_{ij}\right] b_{jkv(t)} & \text{otherwise} \end{cases}$$

3.2.2 Decoding Problem:

In decoding problem of HMM we are to find the most probable path for the sequence of hidden states. While reaching our final state we are given more than one paths to reach out there but the most suitable is the one which is the most probable or in other words the path having higher transition probability and the probability to stay at that state is more than any other state. It says, what is mostly probable sequence ω (t) of hidden states which had led to generation of v(t) that is the visible symbols.

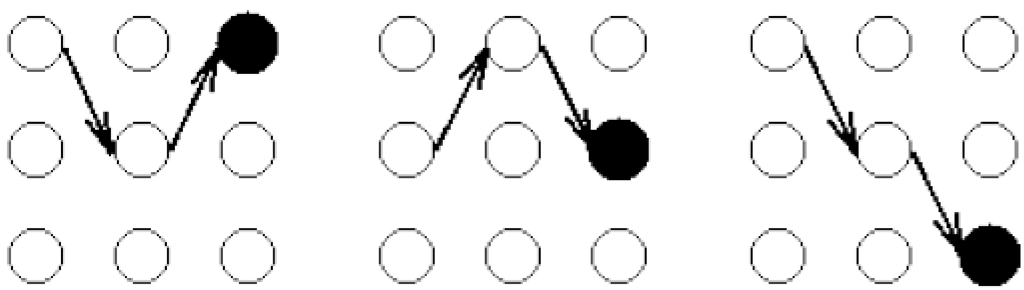


Fig. 3.2 Trellis path

Solution to this problem is the *Viterbi Algorithm* by which you are required to identify the path that is most probable for generating visible sequence of symbols for hidden states. Let's suppose we have hidden states, $\omega_{0,}$ $\omega_{1,}$ $\omega_{2,}$ ω_3 and ω_4 and the visible symbols, v_0, v1, v2, v3 and v_4 And also generating four visible symbols says v^4 = {v_1, $v_{3,}$ v_2 and v_0 }. So, the b_{jk} for this scenario will be;

V_0	V_1	V_2	V_3	V_4	
1	0	0	0	0	ω_0
0	0.3	0.4	0.1	0.2	ω_1
0	0.1	0.1	0.7	0.1	ω_2
0	0.5	0.2	0.1	0.2	ω_3

Table 3.3 Emission probabilities of visible symbols of hidden states

And also the □$_{ij}$ for this scenario will be;

ω_0	ω_1	ω_2	ω_3	ω_4	
1	0	0	0	0	ω_0
0.2	0.3	0.1	0.1	0.4	ω_1
0.2	0.5	0.2	0.7	0.1	ω_2
0.7	0.1	0.1	0.1	0.1	ω_3

Table 3.4 Probabilities of hidden states of Hidden Markov Model

From above table we can easily find the most probable path by looking at the probabilities of emission of visible symbols.

t→	0	1	2	3	4
ω_0	0	0	0	0	**0.00138**
ω_1	1	0.03*0.03= 0.09	0.0052	**0.0019**	0
ω_2	0	0.01	**0.0343**	0.0008	0
ω_3	0	**0.2**	0.0057	0.0012	0
		$\alpha_j(1)$	$\alpha_j(2)$	$\alpha_j(3)$	$\alpha_j(4)$

Table 3.5 Most probable path of sequence for visible symbols of hidden states

So the **P (v^4/ɵ) = 0.00138** that is the probability to reach at hidden state 4 after generating four visible symbols.

3.2.2.1 Viterbi Algorithm:

Initialize: $\quad \text{Path} \leftarrow \{\ \} \quad ; \quad t \leftarrow 0,\ j \leftarrow 0$

for $\quad t \leftarrow t + 1$

$\quad j \leftarrow j + 1$

for $\quad j \leftarrow j + 1$

$$\alpha_j(t) \leftarrow b_{jkv(t)} \sum_{i=1}^{N} \alpha_i\,(t-1).\,a_{ij}$$

until $j = N$

$$j' \leftarrow \arg {}^{\max}{}_{j} \alpha_j\,(t)$$

Append $\omega_{j'}$, to path

until $t = T$

Return Path

3.2.3 Learning Problem:

The hidden and visible states are known, but transition and emission probabilities are to be estimated in this problem associated to HMM. As a solution to this problem a ***backward algorithm or Balm Welch*** is used. It states, identify the probability about model is in state $\omega(j)$ at time instant 't', also produces the rest of sequence of visible states or in other words the expected no. of samples and refined values of transition probabilities starting from t+1 to T.

$$\beta_i(t) = \begin{cases} 0, & \omega_i(t) \neq \omega_0 \quad \text{and } t = T \\ 1, & \omega_i(t) \neq \omega_0 \quad \text{and } t = T \\ \left[\sum \alpha_i (t+1) a_{ij}\right] b_{jkv(t+1)} & \text{otherwise} \end{cases}$$

So the expected number of transitions

$$\omega_i (t-1) \rightarrow \omega_j(t)$$

At time in v^T

$$\sum_{t=1}^{T} \gamma_{ij} \quad (t)$$

Expected number of transitions from ω_i

$$\sum_{t=1}^{T} \sum_{k} \cdot \gamma_{ik} \ (t)$$

So the refined values of transition probability with emission probability will be;

$$a'_{ij} = \frac{\sum_{t=1}^{T} \gamma_{ij}\,(t)}{\sum_{t=1}^{T} \sum_{k} \gamma_{ik}\,(t)} \quad \text{and} \quad b'_{jk} = \frac{\sum_{t=1}^{T} \sum_{l} \gamma_{jl}\,(t) \text{ and for } v(t)=v_k}{\sum_{t=1}^{T} \sum_{l} \gamma_{jl}\,(t)}$$

From above equation in the numerator the considering thing is that whenever the emitting symbol is from v_k and the denominator is irrespective of whether the emitting symbol is from v_k or not. Using these values of alpha and beta we can move towards the estimation of refined values of ϒ.

$$\gamma_{ij} = \frac{\alpha_i(t-1)a_{ij}\, b_{jk}\, \beta_j(t)}{P(^{V^T}/_{\theta})}$$

3.3 Advantages Of Hidden Markov Model:

- Hmm is a considered to be a strong statistical foundation.
- Learning algorithm is efficient as the learning takes place directly from raw sequence data.
- Algorithm allows several treatments of insertion and removal in the form of locally learnable.
- Algorithm is as flexible as it can handle the variables of different lengths.
- Algorithm has found its applications in several areas of research like multiple alignments, data mining, classification purpose, statistical analysis and also the pattern recovery.

Chapter 4

Methodology

4.1 Proposed Methodology:

Principal component analysis (PCA) has an advantage over other techniques that it can be used for dimensionality reduction because processing large amount of data is always a difficult task to do. So the hybrid techniques are applied to process and get some better results. PCA applied at the initial stage to get the principal components PC1 which has the highest information about the data. So, instead of using the whole matrices of images we use only principal components for this purpose. Eigen values were obtained and then the Eigen vectors were identified by finding the highest Eigen values. Those were the basic principal components. Experiments were done on MATLAB 2015a with windows 10 pro. Dimensionality reduction is one of the main tasks in face recognition so that was done at the start of the process.

Feature extraction is to be done with Singular Value Decomposition (SVD), where we are required to get the right and left singular values from the matrices. From the theory of SVD the singular values are always there to represent the whole image instead of using the full matrix. We are required to find where are the most variation occurred and also for different combinations of coefficients. Most of the cases the first or second values of singular matrix represent the most variation or in other words having the most information. These combinations are applied to get the most suitable coefficients for recognition purpose. The main reason behind is, that we are to get the maximum recognition rate with less processing time. If these two tasks operate at the same time then it will be an efficient task otherwise we are to do the compromise between them.

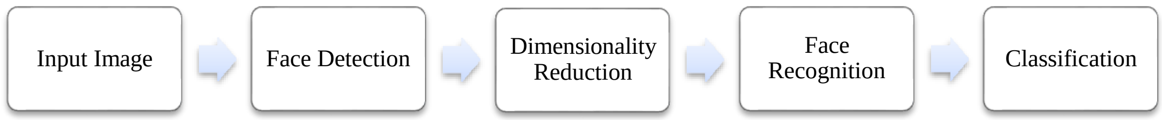


Fig 4.1: Recommended System

In previous techniques only SVD or PCA was applied with some suitable classifier and Euclidean Distance was one of the most prominent but in this task both techniques were applied with proposed classifier for this task i.e. Hidden Markov Model (HMM). It was applied because of its benefits to use, as it the most prominent tool in statistical approaches where you are required to do the classification for images or any sort of data. In this work ORL data set has been used for recognition purpose and that was too for 200 unseen images. PCA with SVD was applied with Euclidean Distance classifier by Kaur and others and their recognition rate was only better for happy that was 95% for Disgust 70% for Surprise 85% for Angry 60 % and for sad 90 % [11]. In this approach for each subject there are 10 images where 5 of them are used for testing and 5 of them are for training images. Image number 1, 5, 7, 8 and 10 are used for training purpose in this approach and the image number 2, 3, 4, 6 and 9 are used for testing purpose. Unrecognized images are shown a cross on them as they were not recognized from training set of images data.

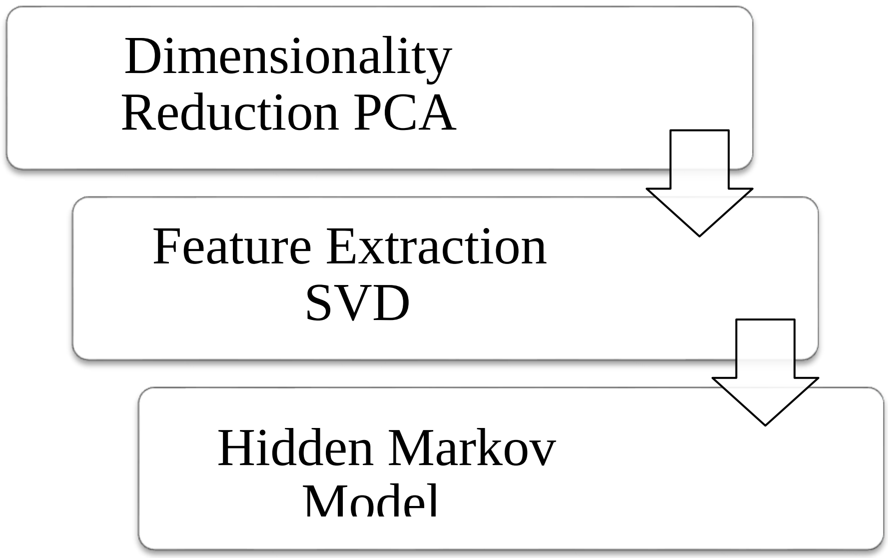


Fig 4.2: Proposed System

In this a comprehensive approach has been applied for recognition of images on ORL database with Hidden Markov Model. Hidden Markov Model (HMM) was also being used as five states model as later it was applied with six and seven states. Eye brows and chin has been added as the other two states and that was efficiently used for classification purposes.

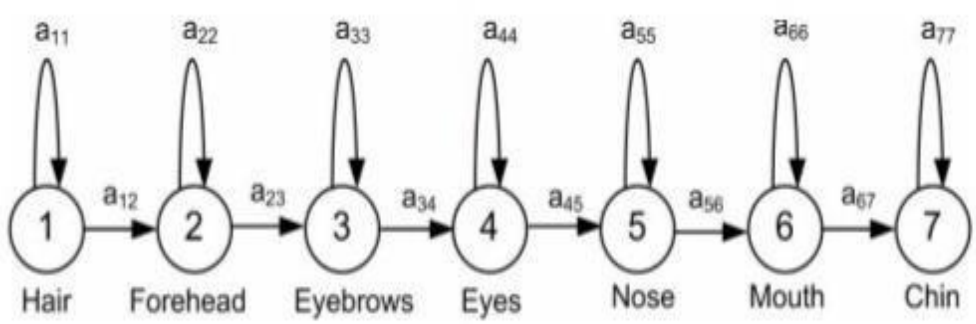


Fig 4.3 Seven states of HMM and their transition probabilities

Fig 4.4 Seven states of HMM for facial image

In this experiment different pairs of coefficients were tried with both the six and seven states of Hidden Markov Model (HMM).

Results were calculated with both the scenarios and better result was obtained with one of the coefficient with seven states model of Hidden Markov Model (HMM).

Peak signal to noise ratio (PSNR) of these recognition rates was obtained as a performance measuring after the outputs of recognition rates. PSNR of 5 training images coefficients was calculated and the result shows its graph.

Chapter 5

Results and Conclusion

5.1 Recognition Rates for 6 states:

Recognition rates were obtained when the Hidden Markov Model was used for its six states model. Following are the recognition rates for different coefficients and their combinations were applied by using Singular Value Decomposition (SVD). These combinations were applied by hit and trial methods and the maximum rate obtained was 86.5%

Coefficients Using 6 States Model	**Recognition Rate**
A(3,3) B(1,1) C(1,1)	69.5%
A(2,2) B(1,1) C(1,1)	73.5%
A(1,2) B(1,1) C(1,1)	74.5%
A(1,1) B(1,1) C(3,3)	85%
A(1,1) B(1,1) C(3,2)	85.5%
A(1,1) B(1,1) C(3,1)	85.5%

A(1,1) B(1,1) C(2,3)	86.5%
A(1,1) B(1,1) C(2,2)	84.5%
A(1,1) B(2,2) C(1,3)	84.5%
A(1,1) B6(2,2) C(1,1)	86%

Table 5.1 Recognition Rates for 6 states

5.2 Recognition Rates for 7 states:

Coefficients Using 7 States Model	**Recognition Rate**
A(1,1) B(2,2) C(1,1)	88%
A(1,1) B(1,1) C(1,2)	84.5%
A(1,1) B(1,1) C(2,2)	84%
A(2,2) B(1,1) C(1,1)	73.5%
A(2,3) B(1,1) C(1,1)	63.5%
A(1,1) B(1,1) C(1,1)	84.5%

Table 5.2 Recognition Rates for 7 states

After using 6 states model I tried these combinations of coefficients by using seven states model and the result was taken better when analyzed and compared with six states model. Maximum recognition rate was obtained using A(1,1) B(2,2) C(1,1) coefficients combinations of SVD and then classified by using HMM for getting the results.

Combinations of PCA and SVD were also applied earlier but their classifier was mostly used Euclidean Distance. So by using HMM we applied a different technique and results were also efficient.

Testing images successfully recognized are shown while other images which are not successfully recognized are shown a cross on them.

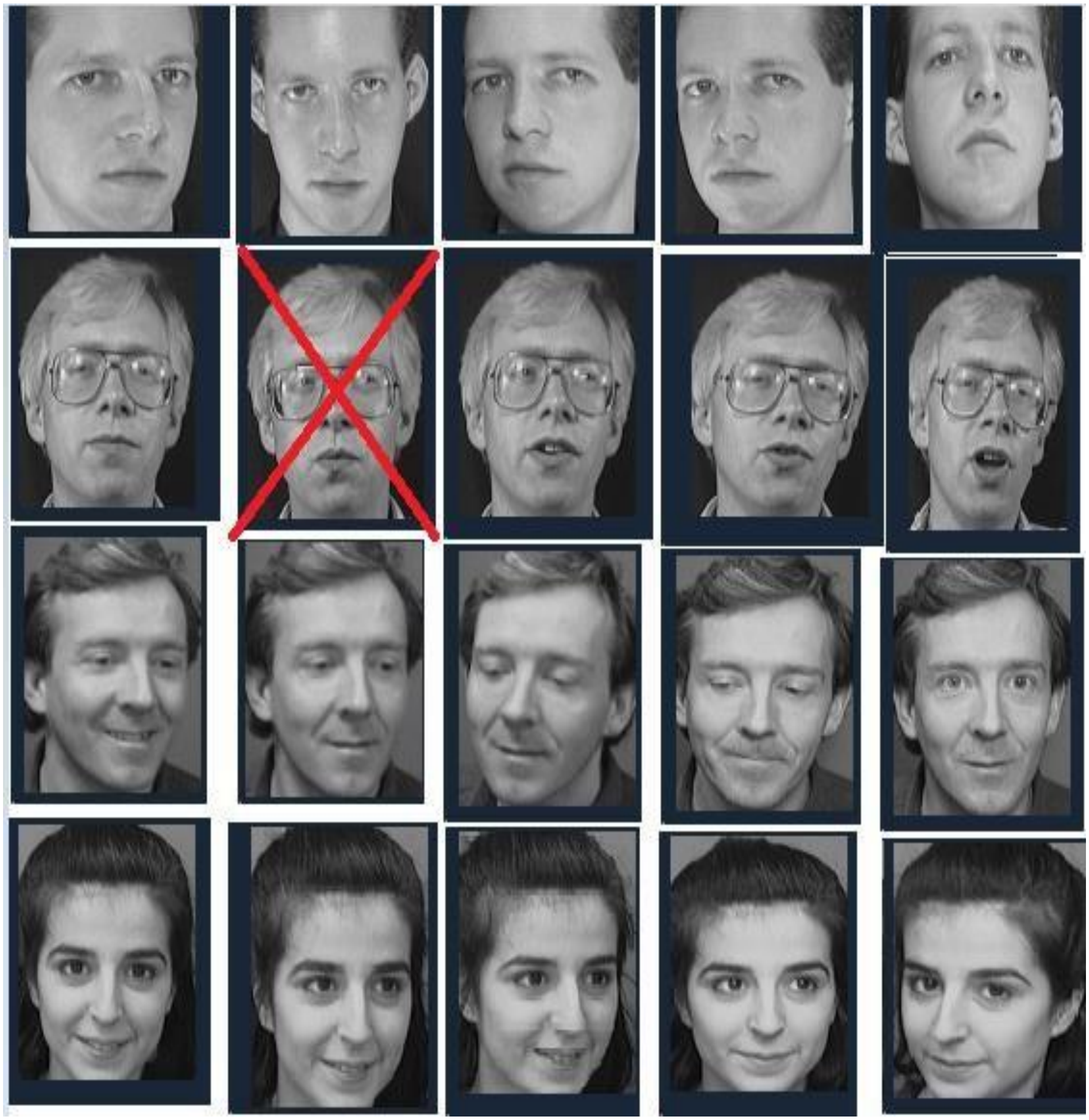


Fig 5.1 Test images successfully/unsuccessfully recognized (a)

Fig 5.2 Test images successfully/unsuccessfully recognized (b)

Images of seven subjects are shown as sample which has been used in our proposed technique. Using this approach satisfactory results have been obtained and also the value of PSNR is also shown in graph.

5.3 PSNR:

Peak signal to noise ratio was calculated and also compared with G.Naresh and Shaik, they applied Euclidean Distance and their recognition rate was 100% when using number count of images was unity which was further reduced when count increased and obtained was 76% for count goes to 10 and the value of PSNR also reduces in this manner [19]. Maximum value in my research work was checked 18 dB in this process. Graph shows the maximum value at E(2b,2b).

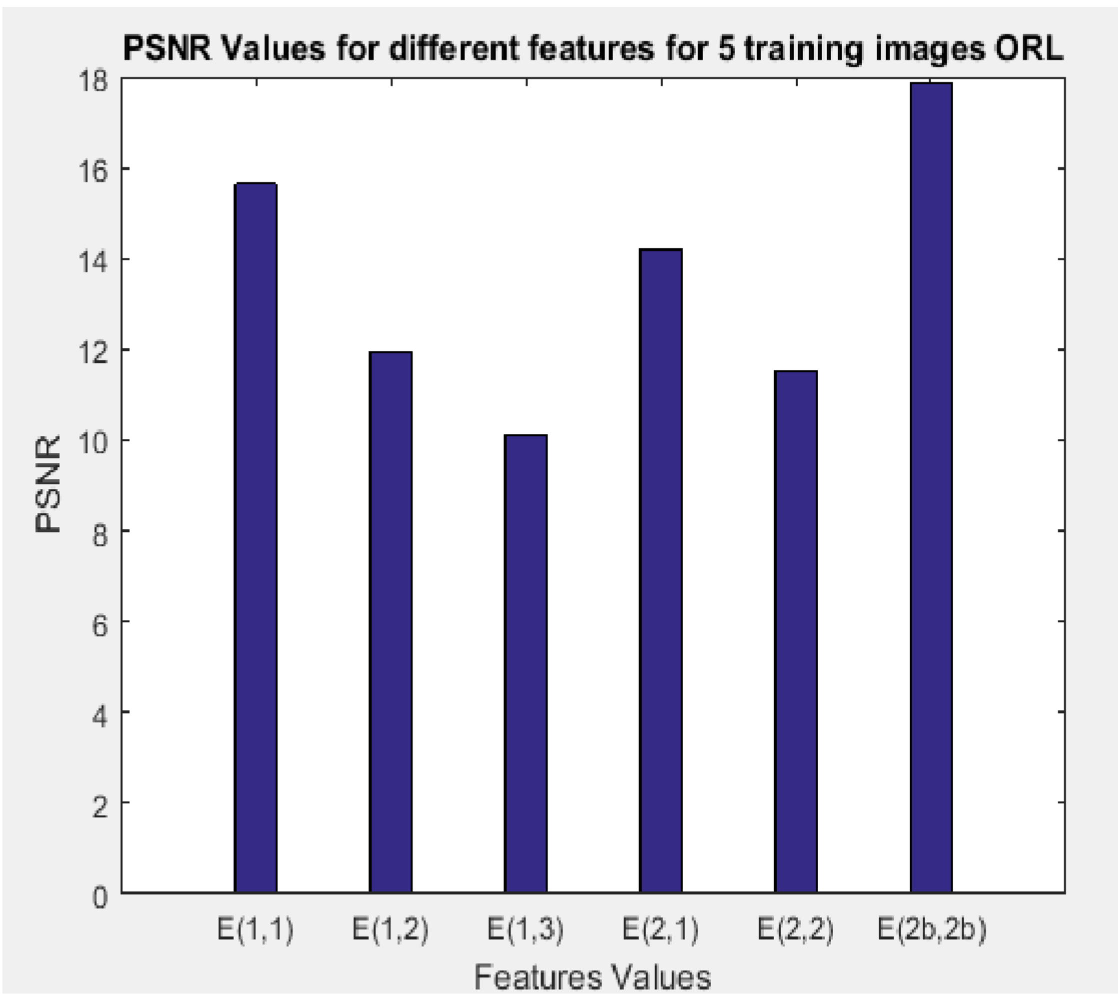


Fig 5.3 PSNR Graph

This combination of techniques was also applied earlier but this was unique for using the classifier of HMM. HMM have some advantages over other classifiers those were discussed earlier in the flow of thesis.

References:

[1] Paul, Liton Chandra, and Abdulla Al Sumam. "Face recognition using principal component analysis method." *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)* 1.9 (2012): 135-139.

[2] Kaur, Mandeep, Rajeev Vashisht, and Nirvair Neeru. "Recognition of facial expressions with principal component analysis and singular value decomposition." *International Journal of Computer Applications* 9.12 (2010): 36-40.

[3] Alaa Eleyan and Hasan Demirel, "PCA and LDA based face recognition using Feed forword neural network classifier", InMultimedia Content Representation, Classification and Security Springer, 2006, pp. 199-206.

[4] Abbas, Eyad I., Mohammed E. Safi, and Khalida S. Rijab. "Face recognition rate using different classifier methods based on PCA." *2017 International Conference on Current Research in Computer Science and Information Technology (ICCIT)*. IEEE, 2017.

[5] Nefian, Ara V., and Monson H. Hayes. "Hidden Markov models for face recognition." *Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP'98 (Cat. No. 98CH36181)*. Vol. 5. IEEE, 1998.

[6] H. Farhan, M. Al-Muifraje, and T. Saeed, "Face recognition using maximum variance and SVD of order statistics with only three states of hidden Markov model," *International Journal of Computer Applications*, vol. 134, pp. 32-39, 2016.

[7] Kaur, Mandeep, Rajeev Vashisht, and Nirvair Neeru. "Recognition of facial expressions with principal component analysis and singular value decomposition." *International Journal of Computer Applications* 9.12 (2010): 36-40.

[8] L. Shen, Z. Ji, L. Bai, and C. Xu, "DWT based HMM for face recognition," *Journal of Electronics (China)*, vol. 24, pp. 835-837, 2007

[9] Samaria, Ferdinando, and Steve Young. "HMM-based architecture for face identification." *Image and vision computing* 12.8 (1994): 537-543.

[10] Bhamre, P. D., and Swati B. Memane. "Face recognition using singular value decomposition and hidden Markov model." *International Journal of Modern Trends in Engineering and Research* 2.10 (2015): 323-332.

[11] Kaur, Mandeep, Rajeev Vashisht, and Nirvair Neeru. "Recognition of facial expressions with principal component analysis and singular value decomposition." *International Journal of Computer Applications* 9.12 (2010): 36-40.

[12] L. Sirovich and M. Kirby. Low-dimensional procedure for the characterization of human faces. J. Opt. Soc. Am. A, 4(3):519–524, Mar 1987.

[13] M. Kirby and L. Sirovich. Application of the karhunen-loeve procedure for the characterization of human faces. IEEE Transactions on Pattern Analysis and Machine Intelligence, 12(1):103–108, Jan 1990.

[14] M. Turk and A. Pentland. Eigenfaces for recognition. Journal of Cognitive Neuroscience, 3 (1):71–86, Jan 1991.

[15] M. A. Turk and A. P. Pentland. Face recognition using eigenfaces. In IEEE Proceedings Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), pages 586–591, Jun 1991.

[16] Bansal, Abhishek, Kapil Mehta, and Sahil Arora. "Face recognition using PCA and LDA algorithm." *2012 Second International Conference on Advanced Computing & Communication Technologies*. IEEE, 2012

[17] Gosavi, Ajit P., and S. R. Khot. "Emotion recognition using principal component analysis with singular value decomposition." *2014 International Conference on Electronics and Communication Systems (ICECS)*. IEEE, 2014.

[18] Sadek, Rowayda A. "SVD based image processing applications: state of the art, contributions and research challenges." *arXiv preprint arXiv:1211.7102* (2012).

[19] G. Naresh and Shaik Rahamtula. An Improvement of Face Recognition By Using Euclidean Distance Method in Dimensionality Reduction Technique. International Journal of Scientific Engineering and Technology Research, pages 1247–1252, March 2015.

[20] R. Kaur and E. Himanshi, "Face recognition using principal component analysis," in *2015 IEEE international advance computing conference (IACC)*, pp. 585–589, IEEE, 2015.

[21] X. Wang, Y. Cai, and M. Abdulghafour, "A comprehensive assessment system to optimize the overlap in dct-hmm for face recognition," in *2015 11th InternationalConference on Innovations in Information Technology (IIT)*, pp. 290–295, IEEE, 2015.