ISSN 2250-3153

Three Fold Security System

Ashwini Sawant*, Bhavesh Vasandani, Mitesh Ochaney**, Mohit Tilwani**, Mohit Kataria****

*Assistant Professor, M.E. EXTC,
** B.E. EXTC students, VESIT, Chembur, Mumbai

Abstract: The paper is divided into following sections namely, text password recognition, text-dependant speech recognition, face recognition. The first stage involves simple alphanumeric character recognition. Mel Frequency Cepstral Coefficients (MFCCs) are the most popularly used speech features in many speech and speaker recognition applications. Speaker identification can be Text-Independent or Text-Dependent. This paper describes an approach of Text-Dependent speaker recognition by using the MFC Coefficients extracted from speech signal of spoken words. The third stage comprises of face recognition for which the proposed method is based on the principle component analysis algorithm. The algorithm MFCC used in this project for speech recognition is the best. This is because it analysis the speech signal on the basis of the human perception of the sound i.e. in the Mel scale. The PCA algorithm is a simple algorithm. This algorithm is used to reduce the computational time of the overall project.

Index Terms*: MFCC, PCA*

I. INTRODUCTION

Speech signal provides information in form of analog signal. It conveys the words and messages being spoken and also provides the identity of the speaker. Speaker recognition is the process of automatically recognizing who is speaking by using the speaker specific information.

Speaker recognition is classified into speaker identification and speaker verification: Speaker identification is the process of determining from which of the registered speakers a given utterance comes. In identification, the number of decision alternatives is dependent on the size of the population. Speaker recognition methods can also be divided into text-dependent (fixed passwords) and text-independent (no specified passwords) methods. This project uses text dependent speaker identification system.

Further, this project implements a face recognition system using the Principal Component Analysis (PCA) algorithm. Automatic face recognition system tries to find the identity of a given face image according to their memory. The memory of a face recognizer is generally simulated by a training set. The training set consists of the features extracted from known face images of different persons. Thus, the task of the face recognizer is to find the most similar feature vector among the training set to the feature vector of a given test image. This project uses PCA as a feature extraction algorithm. This algorithm recognizes the identity of a person where an image of that person (test image) is given to the system.

II. PROPOSED METHOD

A. TEXT PASSWORD

First, the user would be asked to input his or her password which would be a string of minimum 8 alphanumeric characters. Now this password is confirmed and stored in the system. Next time the user logs in, he/she would be required to input his/her password. The password entered would be compared using a string comparison technique. If the password matches, he would be granted access or will be denied.

B. SPEAKER IDENTIFICATION

- **SPEECH PREPROCESSING**

Pre-processing is considered as the first step of speech signal processing, which involve with the analog signal to digital signal conversion [1]. The silence has been removed from the speech signal before any processing has been done on it. The signal is then sampled at a rate of 10000Hz. The speech samples are then segmented into frames of the time length within the range of 20-40msec, also known as Framing. Framing enables the non-stationary speech signal to be segmented into quasi-stationary frames, and enables Fourier Transformation of the speech signal. It is because, speech signal is known to exhibit quasi-stationary behavior within the short time period of 20-40msec. Each individual frame is windowed in order to minimize the signal discontinuities at the beginning and the end of each frame. Here, hamming window is most commonly used as window shape in speech recognition technology which integrates all the closest frequency lines. Impulse response of the Hamming window is shown in the equation below:

$$w(n) = 0.54 - 0.46\cos\left(\frac{2\pi n}{N-1}\right) \quad , 0 \leq n \leq N-1 \ldots\ldots\ldots..(1)$$

$$= 0 \quad , \text{otherwise}$$

Where N is the number of samples in each frame

- **MFCC**

The feature is the spectral envelope of the speech spectrum which is represented by the acoustic vectors. MFCC (Mel Frequency Cepstral Coefficients) is the most common technique for feature extraction which is computed on a warped frequency scale based on known human auditory perception. Based on human perception experiments it is observed that human ear acts as filter i.e.it concentrates on only certain frequency components. Psychophysical studies have shown that human perception of the frequency contents of speech signals does not follow a linear scale. Thus for each tone with an actual frequency, f, measured in Hz, a subjective pitch is measured on a scale called the 'Mel' scale [2], [3]. The Mel-frequency scale has linear frequency spacing below 1000 Hz and a logarithmic spacing above 1000 Hz. Therefore we can use the following approximate formula to compute the Mels for a given frequency f in Hz:

$$Mel(f) = 2595*\log10\left(1+\frac{f}{700}\right) \ldots\ldots\ldots\ldots\ldots\ldots \quad (2)$$

The idea acts as follows as in Figure 1.

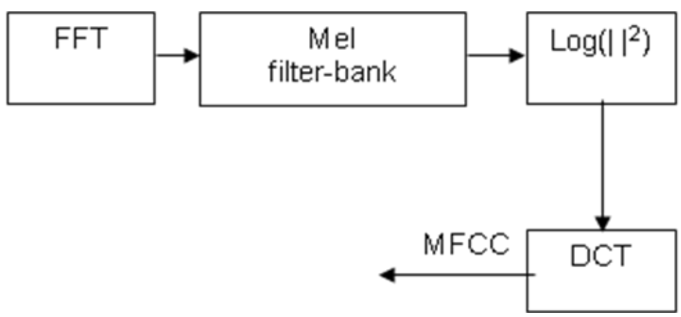


Figure 1. Obtaining Mel cepstrum from windowed frames

Particularly, for the filter banks implementation, the magnitude coefficient of each Fourier Transform speech segment is binned by correlating them with each triangular filter in the filter bank. To perform Mel-scaling, 17 triangular filters having triangular frequency response. The triangular filters are used as they can smoothen the harmonics .There are few filters spaced linearly below 1000 Hz, and the remaining filters spread logarithmically above 1000 Hz. The results of the FFT will be information about the amount of energy at each frequency band. Human hearing, however, is not equally sensitive at all frequency bands. It is less sensitive at higher frequencies, roughly above 1000 Hertz. It turns out that modeling this property of human hearing during feature extraction improves speech recognition performance. Peak (referred to as formants) in the spectrum denote dominant frequency components in the speech signal and carry the identity of the sound. The spectral envelope is used for speaker Identification. We now need to separate spectral envelope and spectral details from spectrum. The next step is to take the logarithm which simply converts the multiplication of the magnitude in the Fourier transform into addition making the extraction of the formants simpler. In general, the human response to the signal level is logarithmic. It is because humans are less sensitive to slight differences in amplitude at high amplitudes compared to the low amplitudes. In addition, using a log makes the feature estimates less sensitive to variations in input (for example, power variations due to the speaker's mouth moving closer or further from the microphone) [4]. The final procedure for the Mel Frequency Cepstral coefficients (MFCC) computation is to convert the log Mel spectrum back to time domain where we get the so called the Mel frequency cepstral coefficients (MFCC). Because the Mel spectrum coefficients are real numbers, we can convert them to the time domain using the Discrete Cosine Transform (DCT) and get a featured vector. The DCT compresses these coefficients to 13 in number. This features vector is considered as an input for the next stage, which are concern with training the features vector and pattern recognition. The cepstrum is more formally defined as the DCT of the log magnitude of the DFT of a signal and is given by:

$$C(u) = a(u) + \sum_{x=0}^{N-1}\left(f(x)\cos\frac{\pi(2x+1)u}{2N}\right) \ldots\ldots\ldots\ldots(3)$$

For u= 0,1,2,...N-1; x=0,1,2,...N-1 and a(u) is defined as -

$$X = (x_1, x_2, \ldots x_r)$$

- **DECISION MAKING ALGORITHM**

The decision is made based on Least Mean Squared Error (MSE). This refers to the mean value of the squared deviations of the predictions from the true values.

$$\textbf{MSE}= \frac{1}{n}\sum(\grave{Y}_i - Y_i)^2 \ldots\ldots\ldots\ldots\ldots\ldots \quad \textbf{(4)}$$

C. FACE RECOGNITION

2.3.1 TRAINING AND TESTING PHASE

In the training phase, feature vectors are extracted for each image in the training set. Let Ω_A be a training image of person A which has a pixel resolution of *M X N* (*M* rows, *N* columns). In order to extract PCA features of Ω_A, we first converted the image into a pixel vector ϕ_A by concatenating each of the *M* rows into a single vector. The length (or, dimensionality) of the vector ϕ_A will be *MXN*. In this project, the PCA algorithm is used as a dimensionality reduction technique which transforms the vector ϕ_A to a vector ω_A which has a dimensionality *d* where $d << M X N$. For each training image Ω_i, feature vectors ω_i were calculated and stored. In the recognition phase (or, testing phase) let *j* be the identity (name) of a person. As in the training phase, feature vector of this person using PCA were computed to obtain ω_j. In order to identify Ω_j, the similarities between ω_j and all of the feature vectors ω_i's in the training set were computed using Euclidean distance. The identity of the most similar ω_i will be the output of the face recognizer. If $i = j$, it means that algorithm correctly identified the person *j*, otherwise if $i \neq j$, it means that the program has misclassified the person *j*.

2.3.2 IMPLEMENTATION OF PCA

In this section, the use of PCA as a feature extractor is explained. Assume that there are p training images: Ω_i; $i = 1, 2 \ldots\ldots p$. For each training image pixel vectors ϕ_i where $\phi_i \varepsilon R^k$; ($k = M X N$) are formed. The aim is to compute feature vectors ω_i where $\omega_i \varepsilon R_d$ $<d$; ($d<<k$). In order to apply PCA to the training set, first a training data matrix *A* is formed which contains *p* rows: at each row ϕ_i's are stored. Thus the dimensionality of A is *pXk*. First, the covariance matrix of *A*: C_A is computed. Then the Eigen values and their responding eigenvectors of C_A are computed. There will be *k* eigen value and eigen vector pairs where each eigenvector e_i is of dimensionality *k*. The eigen values are sorted in decreasing order, and the biggest *d e*igen value and eigenvector pairs is selected. The transformation matrix A is formed by simply putting the selected eigen vectors as columns of A. A is used to compute ω_i's from ϕ_i's .The computation of ω_i is simply done by:

$$\boldsymbol{\omega_i = \Psi^T \phi_i^T} \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad \textbf{(5)}$$

where Ψ^T and ϕ_i^T are the transposes of Ψ and ϕ_i, respectively. Each column of Ψ corresponds to an Eigen vector which is of length *k*. This is equal to *M* x *N* which is the dimensionality (resolution) of input images. Thus, algorithm converted each eigenvector to an image by reversing the concatenation operation. These converted eigenvector images are called Eigen faces since they are similar to human faces.

Once the ω_i's are obtained using the largest *d* eigenvectors, the image of person *i* gets reconstructed. Using all *k* eigenvectors instead of *d* when forming Ψ, the reconstructed image will be the same as image Ω_i. However, since the aim is dimensionality reduction and $d << k$, reconstructed image $\grave{\Omega} i$ will be an approximation of the actual Ωi. You can reconstruct $\grave{\Omega} i$ by converting the pixel vector: $\grave{\Omega}_i = (\Psi\omega_i)^T$ to an image of resolution *MXN*. If more eigenvectors are used, then the reconstructed image is more similar to the original face image.

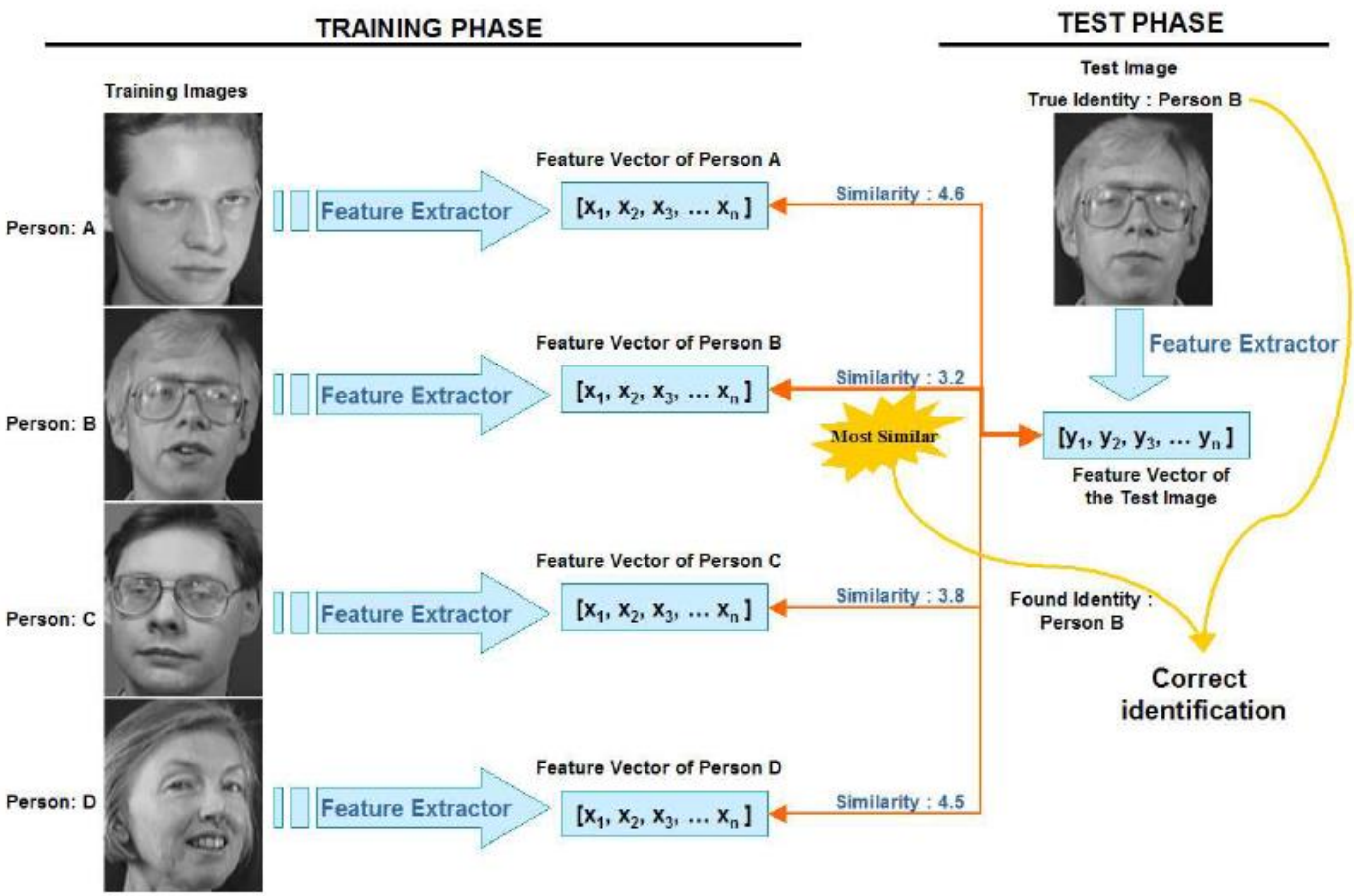


Figure. 2
(Schematic diagram of face recognizer)

III. EXPERIMENTAL RESULT

A. PASSWORD

The Text Password Recognition step looks like:

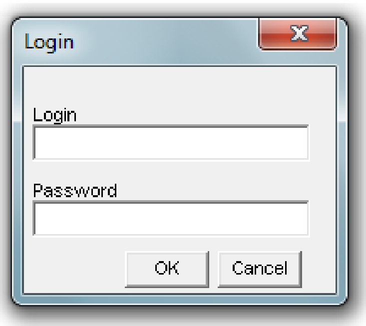


Figure 3. User Interface of Text password

B. SPEECH

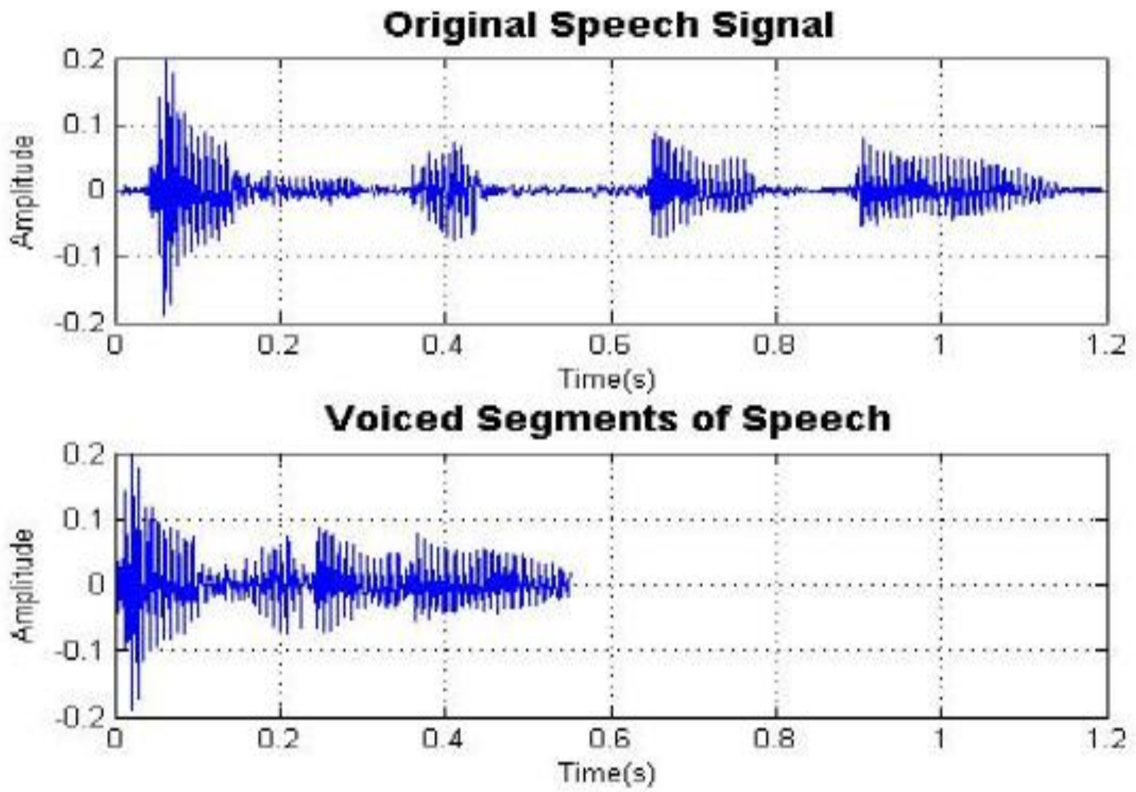


Figure 4. Input and end point detected speech

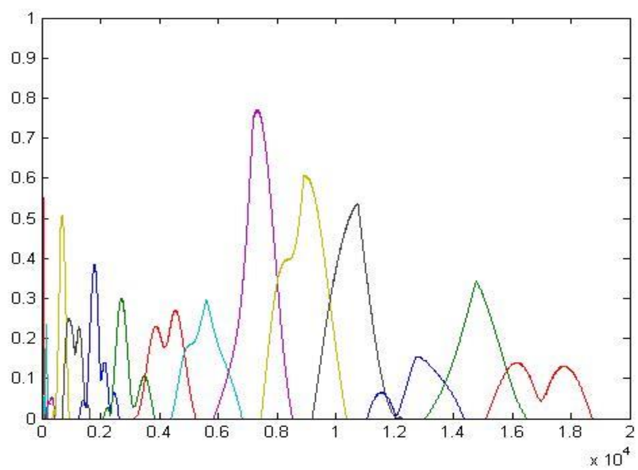


Figure 5. Output of Mel scale filter bank

C. FACE

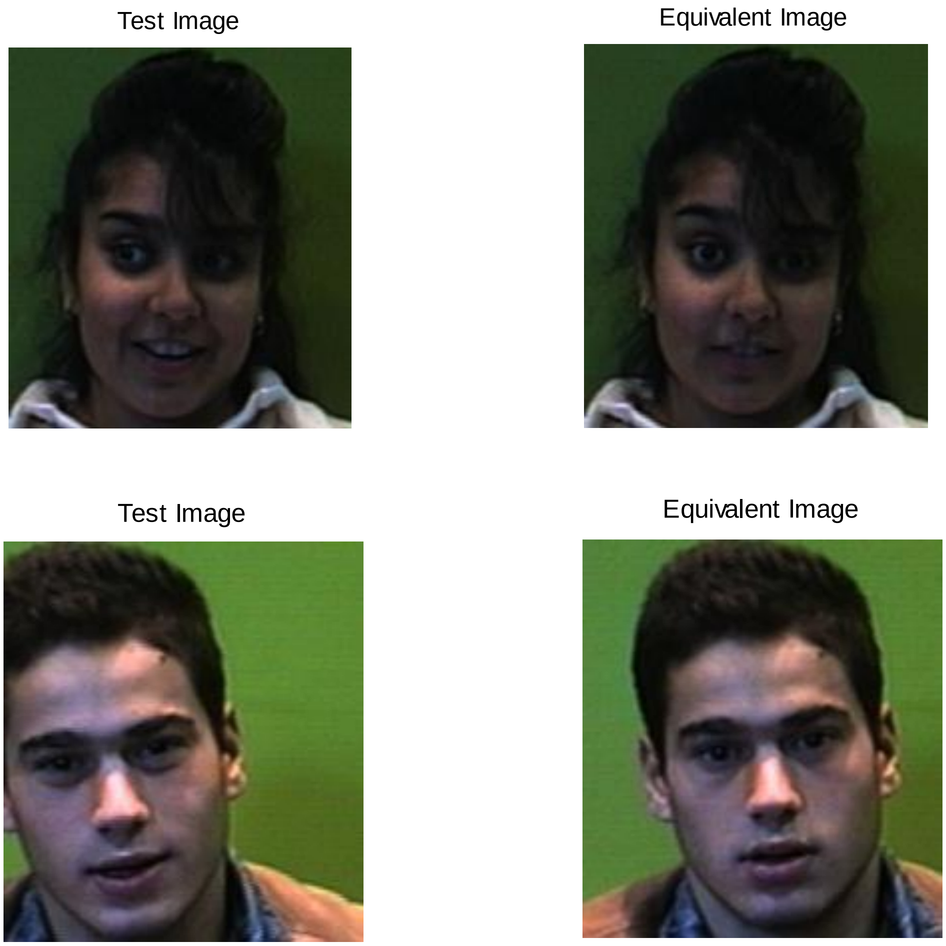
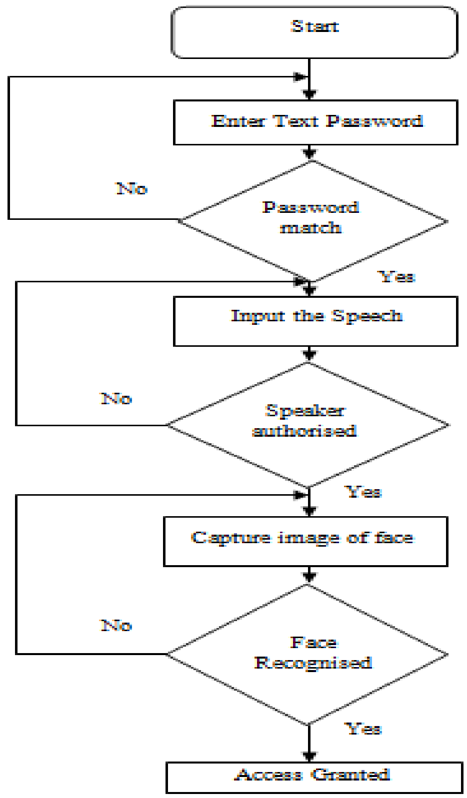


Figure 6. Output of PCA algorithm

IV. CONCLUSION

In this paper, we have presented an approach for a three way Security System allowing access only to authorized users. The text based Password system asks for a login ID and password specific to each user. Further, this paper has evaluated the use of MSE, made using the feature vectors obtained by doing MFCC on the pre-processed speech signal, for Robust Speaker Identification. The results of experiments show that the method is effective. Our project proposes a data security system having three levels of authentication processes which are difficult to surpass, thus avoiding access to those who fail the authentication process. This project can be integrated and used for defense, at entry point of atomic and nuclear research centers, museums, accessing a computer system, an online cloud account which has important documents, online transfer of money for registered user.

V. FLOWCHART OF THE PROJECT



VI. REFERENCES

[1] E.C. Gordon "Signal and Linear System Analysis". Copyright © 1998 John Wiley & Sons Ltd., New York, USA.

[2] S.S. Stevens, J.Volkmann and E.B. Newman, 1937 "A scale for the measurement of the psychological magnitude pitch". Journal of the Acoustical Society of American,8, 185-190.

[3] S.S. Stevens and J.Volkmann, 1940 "The relation of pitch to frequency: A revised scale". The American Journal of Psychology,53(3),329-353.

[4] B. Chanda and D. D. Majumder, Digital Image Processing and Analysis. Prentice-Hall of India Private Limited, 2005.

[5] ORL database available at: http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html.

Authors

First author –Ashwini Sawant, M.E. EXTC, Assistant Professor, VESIT, Chembur and ashwinisawant11@gmail.com.

Second author – Bhavesh Vasandani, B.E. EXTC, Student, VESIT, Chembur and bhavesh_xyzzy@yahoo.co.in .

Third author – Mohit Tilwani, B.E. EXTC, Student, VESIT, Chembur and tilwani.mohit150119@gmail.com.

Fourth author – Mitesh Ochaney, B.E. EXTC, Student, VESIT, Chembur.

Fifth author – Mohit Kataria B.E. EXTC, Student, VESIT, Chembur.