

Dynamic Audio-Visual Client Recognition Modeling

Tijjani Adam

*Institute of Nano Electronic Engineering
Universiti Malaysia Perlis
Kangar, 01000, Malaysia*

tijjaniadam@yahoo.com

U. Hashim

*Institute of Nano Electronic Engineering
Universiti Malaysia Perlis
Kangar, 01000, Malaysia*

uda@unimap.edu.my

Abstract

This paper contains a report on an Visual Client Recognition System using Matlab software which identifies five clients and can be improved to identify as many clients as possible depending on the number of clients it is trained to identify which was successfully implemented. The implementation was accomplished first by visual recognition system implemented using The Principal Component Analysis, Linear Discriminant Analysis and Nearest Neighbor Classifier. A successful implementation of second part was achieved by audio recognition using Mel-Frequency Cepstrum Coefficient, Linear Discriminant Analysis and Nearest Neighbor Classifier the system was tested using images and sounds that have not been trained to the system to see whether it can detect an intruder which lead us to a very successful result with précised response to intruder and also explored another means implementing the visual recognition section using a Neural Network. The work on visual recognition system was converted into a simulink block set which was then implemented in a Signal wave.

Keywords: Audio- visual Client Recognition, Discriminate, Multi-model Biometric System, Simulink, Neural Network.

1. INTRODUCTION

Audio-visual client recognition system is one of the multi-modal biometric systems. The system automatically recognizes or identifies the user based on voice and facial information. The aim of this study is to develop an audio-visual recognition system. The system principal component analysis (PCA) involves a mathematical procedure that transforms a number of possibly correlated variables into a smaller number of uncorrelated variables called principal components[1][2]. The first principal component accounts for as much of the variability in the data as possible, and each succeeding component accounts for as much of the remaining variability as possible. It is used in signal processing for data compression. It can be shown that the best linear compression of a dataset can be achieved by projecting the data onto the Eigen vectors of the data's covariance matrix; the compressed values are thus the principal components of the data. The system is able to identify 5 clients and also detect give an intruder alert. The principal component analysis (PCA) involves a mathematical procedure that transforms a number of possibly correlated variables into a smaller number of uncorrelated variables called principal components [3]. The first principal component accounts for as much of the variability in the data as possible, and each succeeding component accounts for as much of the remaining variability as possible[4]. It is used in signal processing for data compression. It can be shown that the best linear compression of a dataset can be achieved by projecting the data onto the Eigen vectors of the data's covariance matrix; the compressed values are thus the principal components of the data.

After using PCA, unwanted variations caused by the illumination, facial position and facial expression still remain. Accordingly, the features produced by PCA are not necessarily good for Discriminant among classes. Therefore, the most discriminating face features are further acquired by using the LDA method [3][4]. The purpose of LDA is to group images of the same class and separate images of different classes. The nearest neighbour algorithm calculates the distance from the unknown testing sample to every training sample by using the distance metric [5].

Audio Authentication System(AAS): An audio authentication system is a system that identifies a speaker by analyzing spectral shape of the voice signal, usually done by extracting instructions while preparing/modifying these guidelines. This guideline is used for all journals. This guideline is used for all journals. These are the manuscript preparation guidelines used as a standard template for all journal submissions. Author must follow these instructions while preparing/modifying these guidelines. This guideline is used for all journals. This guideline is used for all journals. These are the manuscript preparation guidelines used as a standard template for all journal submissions. Author must follow these instructions while preparing/modifying these guidelines. This guideline is used for all journals and matching the feature of voice signal. Ceptra are most commonly used features used in speech authentication tasks. A Cepstrum of a given signal is obtained using homomorphic filtering which converts a convolved source and filter impulse responses to linear summations. An approach to this is computing the speech linear prediction coefficients (LPCC). An alternative way is to apply a Mel-scale filter-bank function to the speech spectrum. The resulting coefficients are referred to as Mel-Frequency Cepstrum Coefficients (MFCC). There are other types of Ceptra that can be obtained through variations of, or additional processing in, the above methods. Examples of these are perceptual linear prediction coefficients (PLP) and linear filter bank cepstral coefficients (LFCC). LPCC and MFCC are the most widely used speech features. This project focuses only on MFCC because it is the best known and most popular.

The Mel-Frequency Cepstrum Coefficients are coefficients that collectively make up a Mel-Frequency Cepstrum. Mel-Frequency Cepstrum is described as the short-term power spectrum of a sound, based on a linear cosine transform of a log power spectrum on a non linear mel scale frequency[6][7]. The mel scale is a perpetual scale of pitches judged by listeners to be equal in distance from one another. The reference point between this scale and normal frequency measurement is defined by equating a 1000Hz tone, 40dB above the listener's threshold, with a pitch of 1000 mels. Above about 500Hz, larger and larger intervals are judged by listeners to produce equal pitch increments. As a result, four octaves on the hertz scale above 500Hz are judged by listeners to produce equal pitch increments. As a result, four octaves on the hertz scale above 500Hz are judged to comprise about two octaves on the Mel scale. The name Mel comes from the word melody to indicate that the scale is based on pitch comparisons. Its being used is because the human perception of the frequency contents of sound does not follow a linear scale. A neural network (NN) is a network or circuit of biological neurons. A neural network may either be a biological neural network or an artificial neural network[8][9]. Biological neural networks are made up of real biological neurons that are connected or functionally related in the peripheral nervous system or the central nervous system. In the field of neuroscience[10], they are often identified as groups of neurons that perform a specific physiological function in laboratory analysis

1.1 Audio Authentication System

An audio authentication system is a system that identifies a speaker by analyzing spectral shape of the voice signal, usually done by extracting and matching the feature of voice signal[10]. Ceptra are most commonly used features used in speech authentication tasks. A Cepstrum of a given signal is obtained using homomorphic filtering which converts a convolved source and filter impulse responses to linear summations. An approach to this is computing the speech linear prediction coefficients (LPCC). An alternative way is to apply a Mel-scale filter-bank function to the speech spectrum[11]. The resulting coefficients are referred to as Mel-Frequency Cepstrum Coefficients (MFCC). There are other types of Ceptra that can be obtained through variations of, or additional processing in, the above methods. Examples of these are perceptual linear prediction coefficients (PLP) and linear filter bank cepstral coefficients (LFCC). LPCC and MFCC

are the most widely used speech features. This project focuses only on MFCC because it is the best known and most popular. The Mel-Frequency Cepstrum Coefficients are coefficients that collectively make up a Mel-Frequency Cepstrum. Mel-Frequency Cepstrum is described as the short-term power spectrum of a sound, based on a linear cosine transform of a log power spectrum on a non linear mel scale frequency[12]. The mel scale is a perpetual scale of pitches judged by listeners to be equal in distance from one another. The reference point between this scale and normal frequency measurement is defined by equating a 1000Hz tone, 40dB above the listener's threshold, with a pitch of 1000 mels. Above about 500Hz, larger and larger intervals are judged by listeners to produce equal pitch increments[10]. As a result, four octaves on the hertz scale above 500Hz are judged by listeners to produce equal pitch increments. As a result, four octaves on the hertz scale above 500Hz are judged to comprise about two octaves on the mel scale. The name mel comes from the word melody to indicate that the scale is based on pitch comparisons. Its being used is because the human perception of the frequency contents of sound does not follow a linear scale. The formula to convert frequency into mel is $m=2595\log_{10}(f/700+1)=1127\log_{10}(f/700+1)$, Its inverse, f is given by: $f=700 \cdot 10^{m/2595-1}=700 \cdot e^{m/1127-1}$.

The need to recognise individuals is vital to human life. The most natural way to do this is by recognising people's faces or voices. However, it is impossible to personally know everyone that an individual may have to interact with. Biometric devices and technologies automate the process of recognising individuals; Audio-Visual client recognition system is one of the multi-modal biometric systems. The system automatically recognizes or identifies the user based on facial information ranging from few to large number of clients but . The aim of this study is to develop adaptable visual recognition algorithms for client recognition[15,16], the algorithm can be employed for security systems and can be compared to other biometrics such as fingerprint or eye iris recognition systems as well as access control using facial information recognition depending on condition and application is trained for[17,18,19]. The system principal component analysis (PCA) involves a mathematical procedure that transforms a number of possibly correlated variables into a smaller number of uncorrelated variables called principal components[19]. The first principal component accounts for as much of the variability in the data as possible, and each succeeding component accounts for as much of the remaining variability as possible. It is used in signal processing for data compression. It can be shown that the best linear compression of a dataset can be achieved by projecting the data onto the Eigen vectors of the data's covariance matrix; the compressed values are thus the principal components of the data. The system is able to identify 5 clients and also detect give an intruder alert. The principal component analysis (PCA) involves a mathematical procedure that transforms a number of possibly correlated variables into a smaller number of uncorrelated variables called principal components [19]. The first principal component accounts for as much of the variability in the data as possible, and each succeeding component accounts for as much of the remaining variability as possible[18]. It is used in signal processing for data compression. It can be shown that the best linear compression of a dataset can be achieved by projecting the data onto the Eigen vectors of the data's covariance matrix; the compressed values are thus the principal components of the data. After using PCA, unwanted variations caused by the illumination, facial position and facial expression still retain. Accordingly, the features produced by PCA are not necessarily good for Discriminant among classes. Therefore, the most discriminating face features are further acquired by using the LDA method [8,9]. The purpose of LDA is to group images of the same class and separate images of different classes. The nearest neighbour algorithm calculates the distance from the unknown testing sample to every training sample by using the distance metric [10]

1.2 Visual Recognition System Using a Neural Network

A neural network is a network or circuit of biological neurons. A neural network may either be a biological neural network or an artificial neural network. Biological neural networks are made up of real biological neurons that are connected or functionally related in the peripheral nervous system or the central nervous system. In the field of neuroscience, they are often identified as groups of neurons that perform a specific physiological function in laboratory analysis.

Artificial neural networks are made up of interconnecting artificial neurons (programming constructs that mimic the properties of biological neurons). They use mathematical or computational model for information processing based on a connectionistic approach to computations. It is an adaptive system that changes its structure based on external and internal information that flows through the network. They may hence be defined as non-linear statistical modeling or decision making tools. They can be used to model complex relationships between inputs and outputs or to find patterns in data. Artificial neural networks may either be used to gain an understanding of biological neural networks, or for solving artificial intelligence problems without necessarily creating a model of a real biological system. Neural networks [13], as used in artificial intelligence have been viewed as simplified models of neural processing in the brain, even though the relation between this model and brain biological architecture is debated. It is a complex statistical processor. Neural coding is concerned with how sensory and other information is represented in the brain by neurons. The main goal of studying neural coding is to characterize the relationship between stimulus and individual or ensemble neural responses and the relationship among electrical activity of the neurons in the ensemble [14]. It is thought that neurons can encode both digital and analog information. As neural networks emulate human thinking, they are used in visual recognition systems.

2: METHODOLOGY

2.1 The process of Principal Component Analysis

2.1.1 Image data acquisition

Each acquired 2-D facial image was represented by a 1-D vector obtained by concatenating each column/row into a long thin vector in which each vector was represented as

$$X_i = x_{i1} \dots x_{Ni} \dots \dots \dots 2.1.0$$

2.1.2 Data Centering

Each of the training images was mean centered by subtracting the mean image from each of the training images. The mean image is a column vector in which each entry is the mean of all corresponding pixels of the training images in figure 1.

$$X_i = x_i - m \dots \dots \dots 2.1.1$$

$$m = \frac{1}{p} \sum_{i=1}^p x_i \dots \dots \dots 2.1.2$$

2.1.3 Data matrix creation

The vectors were combined side by side and a data matrix of size $N \times P$ was formed, where P is the number of training images and each vector is a single image vector.

$$X = [x_1 \ x_2 \ \dots \ x_p] \dots \dots \dots 2.1.3$$

Covariance was calculated by multiplying the data matrix with its transpose.

$$\hat{U} = XX^T \dots \dots \dots 2.1.4$$

Eigen values and eigenvectors of the covariance were calculated.

$$\Omega V = \hat{U} V \dots \dots \dots 2.1.5$$

The eigenvectors were sorted from high to low according to their corresponding eigenvalues. The eigen vector corresponding to the eigenvector with the largest value is the eigenvector that finds

the greatest variance in the training images. Else, the eigenvector that associated with the smallest Eigen value finds the least variance in the training images.
 An eigenspace V was formed by this eigenvectors matrix.

$$V=V_1V_2\dots V_p \dots\dots\dots 2.1.6$$

2.1.4 Projection of training images

Each of the centred training images was projected into the eigenspace. In this process, the dot product of the centred training image with each of the ordered eigenvectors was calculated.

$$x_i= VTx_i \dots\dots\dots 2.1.7$$

The dot product of the image and the first eigenvector became the first value in the new vector. Steps 1 was carried out in a file “imagepreprocessing.m”, (2.1.1 to 2.1.6) in “pca.m” and (2.1.7) in “TestPCA.m”

2.2 The processes of Linear Discriminant Analysis (LDA)

2.2.1 Within Class scatter matrix calculation

The within class scatter matrix was calculated. It measured the amount of scatter between training images in the same class. It was calculated as the sum of the covariance matrices of the centred images in the i th class.

$$S_i=x_i(x_i-m_i)(x_i-m_i)^T \dots\dots\dots 2.2.1$$

$$S_W=\sum_{i=1}^C S_i \dots\dots\dots 2.2.2$$

Where m_i : is the mean of training images within the class. The within class scatter matrix (S_W) is the sum of all the scatter matrices.

The class scattered matrix (S_B) was calculated as the sum of the covariance matrices of the difference between the total mean and the mean value in each class.

$$S_B=\sum_{i=1}^C C_n(m_i-m)(m_i-m)^T \dots\dots\dots 2.2.3$$

The generalized eigenvectors (V) and eigenvalues (L) of the within class and between class scatter matrices were computed.

$$S_B=S_WV \dots\dots\dots 2.2.4$$

The non-zero eigenvectors were sorted from high to low according to their corresponding eigenvalues and the first $C-1$ eigenvectors were kept which formed the Fisher basis vector. The training images were projected onto the fisher basis vector by calculating the dot product of the training image with each of the Fisher basis vector. The above steps were carried out in a file “lda.m”.

2.3 Audio Authentication System Process

The system was designed in such a way to be trained with sound clips each of which represents the identity of an individual. After it has been trained, the system receives an input signal, compare against all possible combinations of sounds it was trained with and then authenticates whether the inputted sound matches any of the sound it was trained with or not. If there is a match it welcomes a particular individual. In the absence of any match it produces an intruder alert. The training was carried out with five different sound clips representing the identity of a single individual so that it will be familiar with each person’s voice and it was done with following steps; **Data Extraction:** data stored in audio wave format is converted to a form suitable for

further computer processing and analysis, **Data Pre-processing**: removing pauses, silences, weak unvoiced sound signals and valid data detection. A digital was defined to perform this task and the related programmed codes were saved as “vad.m”, **Mel-Frequency Cepstrum**: convert the speech waveform to some type of parametric representation (at a considerably lower information rate) for further analysis and processing. This is referred to as signal processing front end, **Frame Blocking**: The continuous speech signal is blocked into frames of N samples, with adjacent frames being separated by M ($M < N$). The first frames consist of N samples. The second frame begins on samples after the first frame, and overlaps it by N-M samples. This process continues until all the speech is accounted for one or more frames, **Windowing**: spectrum distortion minimization by using a window to taper the signal to zero at the beginning and end of each frame. If the window is $w(n)$, $0 \leq n \leq N-1$, result; $y_1n=x_1nw_n$, $0 \leq n \leq N-1$, **Fast Fourier Transform (FFT)**: $X_n=K=0N-1x_k e^{-2\pi jknN}$, $n = 0, 1, 2, \dots, N-1$, **Mel-Frequency Wrapping**: The spacing and bandwidth of such a filter is determined by a constant Mel-frequency interval. The modified spectrum of $S(w)$ consists an output power of these filters when $S(w)$ is the input. The number of Mel-spectrum coefficients, K was chosen to be 13, **Cepstrum**: The cepstral representation of the speech spectrum provides a good representation of the local spectral properties of the signal for the given frame analysis. Because the Mel-spectrum coefficients (and their logarithm) are real numbers, they are being converted to the time domain using Discrete Cosine Transform (DCT). In this step the Mel-Frequency Cepstrum Coefficients are finally found. The set of coefficients is called an acoustic vector. Hence each input utterance is transformed into a sequence of acoustic vectors, **Feature Matching**: Previous Nearest Neighbour algorithm was used as in the case of the Visual recognition system.

2.4 : Visual Recognition System (Neural Network) Process

The input to neural network visual recognition system is a digitized image in matrix form, **Learning process**; The neural networks function is to distinguish between different inputs fed to it and identify each one correctly. It does this by first learning the various images and their classes. In this project the main area of concern is the human visual recognition system. An individual has different moods. Images of an individual in different moods will be recognized by the system differently. This is the reasons why classes are needed and a class is just a collection of images of an individual in different moods. Each image in a class can represent the individual's identity. Images are taught to the network. The input to the network is an input matrix, P. Any neural network may learn in a supervised or unsupervised manner by adjusting its weight. Each individual image taught to the network possesses a corresponding weight matrix. For the nth image taught to the network, the weight matrix is denoted by W_n . As the learning process continues, the weight matrix is being upgraded. Each image inputted into the neural network is stored under a class. The weight matrix is updated as shown below:

for all $i = 1$ to x

{

for all $j = 1$ to y

{ $W_n(i, j) = W_n(i, j) + P(i, j)$

}

}

x and y are the dimensions of matrix W_n and P . **Candidate Score (\square)**: This is the product of corresponding elements of the weight matrix W_n of the n^{th} learnt image and an input image as its candidate. It is obtained using the equation

$\square = \sum_{i=1}^x \sum_{j=1}^y W_{ki,j} * P(i,j)$.

Ideal Weight Model Score (μ): This gives the sum total of all the positive elements of the weight matrix of a learnt image. It can be formulated as follows;

For $i = 1$ to x

{

For $j = 1$ to y

```

{
If WK (i, j) > 0 then
{
 $\mu(K) = \mu(K + WK(i, j))$ 
}
}
}
}

```

Similar to the visual recognition system using PCA, LDA and nearest neighbour algorithm in parts 1 to 3, training images 1 to 5 can be taught to the network as a single class representing Mr. A. **Recognition Quotient (Q):** This statistics gives a measure of how well the recognition system identifies an input image as a matching candidate for one of its many learnt images. It is $Q_k = \frac{1}{\mu(K)} \sum \mu(K)$. The greater the value of Q, the more accurate the system is in identifying individuals. A low value of Q indicates a poor recognition. In this case the individual does not exist within the knowledge base or the network has to be taught until a satisfactory Q is obtained.

2.5 Implementation Using Matlab Software

This involved the use of the Neural Network Toolbox. To define an image recognition problem, there must be a set of *Q* input vectors as columns in a matrix. Another set of *R* target vectors have to be arranged so that they indicate the classes to which the input vectors are assigned. The target vector must have the following properties

1. Row headers – column vector containing names of all row headers.
2. Col headers – row vector containing the names of all column headers.
3. Textdata - matrix containing all imported test data. Empty elements are set to „□“.
4. Data – matrix containing all imported numeric data. Empty elements are set to NaN

With the neural network toolbox, previously in part1 the training images were uploaded into workspace in a file “*imagepreprocessing.m*”. As a necessity to arrange the training images as columns in a vector, it has been already done in the file and so the generated matrix was used as the input vector. A target vector was then generated in M-file by name “*target.m*” in which images were assigned the following target values

TABLE1: Target values of training images. Images that represent the same individual were assigned the same target values.

N0.	Image	Target Value
1	1-5	001
2	6-10	010
3	11-15	011
4	16-20	100
5	21-25	101

After the development of simulink block from command “nprtool” to the finish using Matlab command window, the Visual Recognition system was converted into simulink block set as shown in figure2. This involved calculations of eigenvectors, which makes it tedious to implement using simulink. As such the results of eigenvector calculations will be imported into simulink and will show the identity of each individual by number

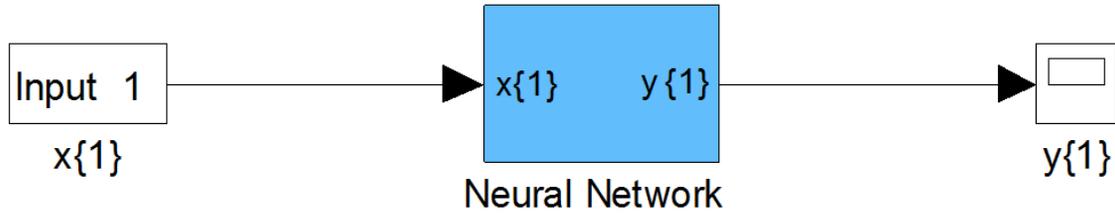


Figure2: Simulink Block set of the Neural Network, The block Neural Network is a subsystem with many components in it.



FIGURE1: Training images and sound used in this study

3: RESULTS AND DISCUSSION

The model feature is summarized as follow: the extracted facial signals and template representing each individual is constructed, there remains the task of classifying live samples. In the case of identity verification the live sample features must be compared with the template associated with the claimed identity. The distance between these two points in the feature space will have to be compared with a threshold of acceptability. This threshold is set empirically to produce acceptable performance. In the case of person identification the live sample will have to be compared with all the stored templates and a range of distances will be measured. The closest match with the smallest distance will then be chosen as the identity. Various architectures have been used for performing such classifications. There is usually a training phase where the classifier is given valid feature vectors and their associated identity tokens. The success of the operational phase depends on the quality of this training phase. The system can be arranged so that if the any of the modalities produce an acceptance then the user is accepted and the other

layers need not be invoked. It is also possible to have a logical operation performed at the final stage to combine the decisions at the parallel layers.

In a layered approach several modalities of biometric information may be integrated. The nearest neighbour algorithm calculates the distance from the unknown testing sample to every training sample by using the distance metric [9]. Euclidean distance was used in the distance computation. It is the last stage in visual recognition system. The mathematics formula of Euclidean distance is:

$$D_{xi, yj} = \sqrt{\sum_{j=1}^n (x_j - y_j)^2} \quad (3-1)$$

The MATLAB codes showing the above performed task are contained in a file "PCALDAnn.m". The results below were obtained when the main program was run (PCALDAmainprog.m). They were obtained when the testing image directory was set as the testing database (imdir='E:\DSP Project Files\Testingdatabase10imgs\');. We shown a set of results also when the fraud database is set as the testing image directory in figure 1.

3.1 Data inputting

```
Input =  
E:\DSP Project Files\Trainingdatabase25imgs\1.bmp  
Input =  
E:\DSP Project Files\Trainingdatabase25imgs\2.bmp  
Input =  
E:\DSP Project Files\Trainingdatabase25imgs\3.bmp  
Input =  
E:\DSP Project Files\Trainingdatabase25imgs\4.bmp  
Input =  
E:\DSP Project Files\Trainingdatabase25imgs\5.bmp  
Input =  
E:\DSP Project Files\Trainingdatabase25imgs\6.bmp  
Input =  
E:\DSP Project Files\Trainingdatabase25imgs\7.bmp  
Input =  
E:\DSP Project Files\Trainingdatabase25imgs\8.bmp  
Input =  
E:\DSP Project Files\Trainingdatabase25imgs\9.bmp  
Input =  
E:\DSP Project Files\Trainingdatabase25imgs\10.bmp  
Input =  
E:\DSP Project Files\Trainingdatabase25imgs\11.bmp  
Input =  
E:\DSP Project Files\Trainingdatabase25imgs\12.bmp  
Input =  
E:\DSP Project Files\Trainingdatabase25imgs\13.bmp  
Input =  
E:\DSP Project Files\Trainingdatabase25imgs\14.bmp  
Input=  
E:\DSP Project Files\Trainingdatabase25imgs\15.bmp  
Input =  
E:\DSP Project Files\Trainingdatabase25imgs\16.bmp  
Input =  
E:\DSP Project Files\Trainingdatabase25imgs\17.bmp  
Input =  
E:\DSP Project Files\Trainingdatabase25imgs\18.bmp  
Input=  
E:\DSP Project Files\Trainingdatabase25imgs\19.bmp
```

Input =
E:\DSP Project Files\Trainingdatabase25imgs\20.bmp
Input =
E:\DSP Project Files\Trainingdatabase25imgs\21.bmp
Input =
E:\DSP Project Files\Trainingdatabase25imgs\22.bmp
Input =
E:\DSP Project Files\Trainingdatabase25imgs\23.bmp
Input =
E:\DSP Project Files\Trainingdatabase25imgs\24.bmp
Input =
E:\DSP Project Files\Trainingdatabase25imgs\25.bmp

3.2 Test Input

Testinput =
E:\DSP Project Files\Testingdatabase10imgs\ppl1.bmp
Testinput =
E:\DSP Project Files\Testingdatabase10imgs\ppl2.bmp
Testinput =
E:\DSP Project Files\Testingdatabase10imgs\ppl3.bmp
Testinput =
E:\DSP Project Files\Testingdatabase10imgs\ppl4.bmp
Testinput =
E:\DSP Project Files\Testingdatabase10imgs\ppl5.bmp
Testinput =
E:\DSP Project Files\Testingdatabase10imgs\ppl6.bmp
Testinput =
E:\DSP Project Files\Testingdatabase10imgs\ppl7.bmp
Testinput =
E:\DSP Project Files\Testingdatabase10imgs\ppl8.bmp
Testinput =
E:\DSP Project Files\Testingdatabase10imgs\ppl9.bmp
Testinput =
E:\DSP Project Files\Testingdatabase10imgs\ppl10.bmp

3.3 Simulation Results

Ans=
Welcome! Ms. A
Ans =
Welcome! Ms. A
Ans =
Welcome! Ms. B
Ans =
Welcome! Ms. B
Ans =
Welcome! Ms. C
Ans =
Welcome! Ms. C
Ans =
Welcome! Ms. D
Ans =
Welcome! Ms. D
Ans =
Welcome! Ms. E
Ans =
Welcome! Ms. E

3.4 Tests for Extruder

RESULTS OBTAINED USING THE FRAUD DATABASE AS TESTING DATABASE

In this case, the testing database was set to be imdir='E:\Fraud\DSP Project Files\Fraud database\'. The results obtained are shown below:

3.5 Input Data

Input =
E:\DSP Project Files\Trainingdatabase25imgs\1.bmp
Input =
E:\DSP Project Files\Trainingdatabase25imgs\2.bmp
Input =
E:\DSP Project Files\Trainingdatabase25imgs\3.bmp
Input=
E:\DSP Project Files\Trainingdatabase25imgs\4.bmp
Input =
E:\DSP Project Files\Trainingdatabase25imgs\5.bmp
Input =
E:\DSP Project Files\Trainingdatabase25imgs\6.bmp
Input =
E:\DSP Project Files\Trainingdatabase25imgs\7.bmp
input =
E:\DSP Project Files\Trainingdatabase25imgs\8.bmp
input =
E:\DSP Project Files\Trainingdatabase25imgs\9.bmp
input =
E:\DSP Project Files\Trainingdatabase25imgs\10.bmp
input =
E:\DSP Project Files\Trainingdatabase25imgs\11.bmp
input =
E:\DSP Project Files\Trainingdatabase25imgs\12.bmp
input =
E:\DSP Project Files\Trainingdatabase25imgs\13.bmp
input =
E:\DSP Project Files\Trainingdatabase25imgs\14.bmp
input =
E:\DSP Project Files\Trainingdatabase25imgs\15.bmp
Input =
E:\DSP Project Files\Trainingdatabase25imgs\16.bmp
Input =
E:\DSP Project Files\Trainingdatabase25imgs\17.bmp
Input =
E:\DSP Project Files\Trainingdatabase25imgs\18.bmp
Input =
E:\DSP Project Files\Trainingdatabase25imgs\19.bmp
Input =
E:\DSP Project Files\Trainingdatabase25imgs\20.bmp
Input=
E:\DSP Project Files\Trainingdatabase25imgs\21.bmp
Input =
E:\DSP Project Files\Trainingdatabase25imgs\22.bmp
Input =
E:\ DSP Project Files\Trainingdatabase25imgs\23.bmp
Input =
E:\DSP Project Files\Trainingdatabase25imgs\24.bmp

Input =
E:\DSP Project Files\Trainingdatabase25imgs\25.bmp

3.6 Testinput =

E:\Fraud\ DSP Project Files\Fraud database\ppl1.bmp

Testinput =

E:\Fraud\ DSP Project Files\Fraud database\ppl2.bmp

Testinput =

E:\Fraud\ DSP Project Files\Fraud database\ppl3.bmp

Testinput =

E:\Fraud\ DSP Project Files\Fraud database\ppl4.bmp

Testinput =

E:\Fraud\ DSP Project Files\Fraud database\ppl5.bmp

3.7 Simulation

Ans =

Get Away Intruder!!

3.8 Simulation results on Audio Authentication System

ans =

Welcome!Ms. A

ans =

Welcome!Mr. B

ans =

Welcome!Mr. B

ans =

Welcome!Ms. C

ans =

Get Away Intruder!!

ans =

Welcome! Mr. D

ans =

Welcome!Mr. D

ans =

Welcome!Mr. E

ans =

Welcome!Mr. E

Simulation results showed that the system was able to identify the clients. It identified the second testing voice of Ms. C as an intruder. The source voices were listened to and it was observed that voice from the source file has the farthest difference from the training ones compared to the other test voice. This is an indicator that the system may require more training for the voice section

which can be accomplished by increasing the number of training images representing a single identity to say ten. As for detecting a fraud, the system has given a full assurance that it will reject any fraud voice

3.9 Implementation Using Simulink

In the previous program, each time an individual has been identified, the identity matrix is updated. As such, the program was modified in such a way that the identity matrix will show the identity of each and every individual by column. The number of column is ten since we have ten test pictures. This was uploaded into simulink and the simulink was run after completing. the following results were obtained.

```
val(:,:,1) = 4  
val(:,:,2) =4  
val(:,:,3) =8  
val(:,:,4) =6  
val(:,:,5) =12  
val(:,:,6) =15  
val(:,:,7) = 17  
val(:,:,8) =18  
val(:,:,9) =21  
val(:,:,10) =23
```

Results showed the identity numbers of the test images. In file "PCALDAnn.m", the coding was done in such a way that identity numbers 1-5 represent Ms. A, 6-10 for Mr. B, 11-15 for Ms. C 16-20 Mr D, and 21-25 Mr E. If we compare this results with the ones obtained previously above it will be observed that the system was able to identify each individual correctly.

4: SUMMARY AND FIELDS OF APPLICATION

The implementation was accomplished first by visual recognition system implemented using The Principal Component Analysis, Linear Discriminant Analysis and Nearest Neighbour Classifier. The system was tested using images that have not been trained to the system to see whether it can detect an intruder which led us to a very successful result with precised response to intruder. For verification purpose, we explored another means of implementing the visual recognition system using a Neural Network, the work on visual recognition system was converted into a simulink block set which was then implemented in a Signal wave and the system was able to identify each individual correctly. For biometric system in its identification mode may be deployed to monitor surveillance cameras and/or the telephone system within a school campus or residential to identify known specific individuals who may have been excluded from parts or all of the facilities in compound. These could be known debtors, troublemakers, shoplifters etc. In this mode the system will have been supplied with template information for specific individuals and will continuously search for a match or difference with the facial information. Although tremendous efforts were made in the field but most of this system is customized and not affordable by general public, it is the effort of the study to generate very simple algorithm which has a potential of adaptability to many different conditions and could be simply interfaced with many electromechanical or otherwise through very simple digital interfacing

5: CONCLUSION

The results showed the accuracy of the visual recognition system in identifying test images tested against the trained images. When the system was tested with images it wasn't trained with, it was able to give a feedback that an intruder was detected. Thus the PCA, LDA and Nearest Neighbor algorithm are powerful tools in developing such systems. By increasing the number of training images as well as adjusting the parameters picnum and ppl the system can be upgraded to identify an unlimited number of clients, audio recognition was implemented and second voice C is identified as extruder and system has given a full assurance that it will reject any fraud voice

6: ACKNOWLEDGEMENT

The authors wish to thank Universiti Malaysia Perlis (UniMAP), Ministry of Science, Technology & Innovation (MOSTI) and Ministry of Higher Education for giving the opportunities to do this research in the Micro & Nano simulation and Fabrication Lab for providing us a system to generate the codes. The appreciation also goes to all the team members in the Institute of Nanoelectronic Engineering especially in the Nano Biochip Research Group and electronics communication research group for useful discussions.

7: REFERENCE

- [1] Raychaudhuri, S., Stuart, J.M. and Altman, R.B. Principal components analysis to summarize microarray experiments: application to sporulation time series. Pacific Symposium on Biocomputing (2000).
- [2] Jonathon Shlens, A Tutorial on Principal Component Analysis Center for Neural Science, New York University New York City, NY 10003-6603 and Systems Neurobiology Laboratory, Salk Institute for Biological Studies La Jolla, CA 92037(2009)
- [3] S. Gong et al., Dynamic Vision: from Images to Face Recognition, Imperial College Press, 2001 (pp. 168-173 and Appendix C: Mathematical Details, hard copy).
- [4] H. Murase and S. Nayar, "Visual Learning and Recognition of 3D Objects from Appearance", *International Journal of Computer Vision*, vol 14, pp. 5-24, 1995 (hard-copy)
- [5] D. L. Swets and J. Y. Weng. Using discriminant eigenfeatures for image retrieval. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 18(8):831–836, 1996.
- [6] W. Liu, Y. Wang, S. Z. Li, and T. Tan. Null space approach of Fisher discriminant analysis for face recognition. In Proc. European Conference on Computer Vision, Biometric Authentication Workshop, 2004.
- [7] Z. Jin, J. Y. Yang, Z. S. Hu, and Z. Lou. Face recognition based on the uncorrelated Discriminant transformation. *Pattern Recognition*, 34:1405–1416, 2001
- [8] M. A. Turk and A. P. Pentland. Face recognition using Eigenfaces. In Proc. Computer Vision and Pattern Recognition Conference, pages 586–591, 1991.
- [9] KAI YU, LIANG JI* and XUEGONG ZHANG, Kernel Nearest-Neighbor Algorithm, *Neural Processing Letters* 15: 147^156, 2002 Tsinghua University, Beijing, P.R. China, 100084; tel: 86-10-62782877 fax: 86-10-62784047)
- [10] Stéphane Dupont and Juergen Luetttin, Audio-Visual Speech Modeling for Continuous Speech Recognition, *Eee transactions on multimedia*, vol. 2, no. 3, september 2000.
- [11] E. Martinian, S. Yekhanin, and J. Yedidia, "Secure biometrics via syndromes," in Proc. Allerton Conf. Commun., Contr. And Comput., Allerton, IL, 2005.
- [12] D. Varodayan, Y.-C. Lin, A. Mavlankar, M. Flierl, and B. Girod, "Wyner-Ziv coding of stereo images with unsupervised learning of disparity," in Proc. Picture Coding Symp. Lisbon, Portugal, 2007.
- [13.] Abu-Mostafa, Y. (1993), "Hints and the VC Dimension", *Neural Computation*, Vol. 5, No. 2, pp. 278–288.

- [14] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models—their training and application," *Comput. Vis. Image Understand.*, vol. 61, pp. 38–59, Jan. 1995.
- [15] J R Parks, "Automated Personal Identification Methods for Use with Smart Cards", Chapter 7 in *Integrated Circuit Cards, Tags and Tokens* edited by P Hawkes, D Davies and W Price, BSP, London, ISBN 0-632-01935-2, 1990
- [16] F Deravi, "Audio-Visual Person Recognition for Security and Access Control", JTAP project JTAP-552, local web site, URL: <http://eleceng.ukc.ac.uk/~fd4/jtap.html>, 1998-1999
- [17] L Burbridge, Experience with the use of a multi-purpose smart card, JTAP Report 019, JISC, March 1998.
- [18] C C Chibelushi, J S D Mason and F Deravi, "Feature-level Data Fusion for Bimodal Person Recognition", Sixth International Conference on Image Processing and its Applications, IEE, Trinity College, Dublin, Ireland, , pp 339-403, 14-17 July, 1997
- [19] J D Woodward, "Biometrics: Privacy's Foe or Privacy's Friend?", Proceedings of the IEEE, Vol.85, No. 9, September 1997.