International Journal of Current Engineering and Technology
E-ISSN 2277 – 4106, P-ISSN 2347 - 5161

Available at http://inpressco.com/category/ijcet

Research Article

A Security System by using Face and Speech Detection

Chandrashekhar.S.Patil[Ȧ] and Gopal.N.Dhoot[Ȧ*]

[Ȧ]EXTC,Engg.Department,Shri Gulabrao Deokar College of Engineering,Jalgaon,Maharashtra,India.

Accepted 10 June 2014, Available online 28 June2014, **Vol.4, No.3 (June 2014)**

Abstract

The multimodal biometric system for identity verification using two traits i.e. face and speech. The proposed system is designed for applications where the training database contains a face and speech. The final decision is made by fusion at matching score level in which feature vectors are created independently for query images and are then compared to the enrollment templates which are stored during database preparation for each biometric trait. Based on the proximity of feature vector and template, each subsystem computes its own matching score. These individual scores are finally combined into a total score, which is passed to the decision module. Multimodal system is developed through fusion of face and speech. This system is tested on ORL database and the overall accuracy of the system is found by doing experiments yielded as 91.25% for face recognition while recognition rates for speakers 77.78 %. However, the final recognition decision for authorization or access activation is based on the recognition outcomes of the face and speech detection.

Keywords:*Face,Speech,Detection,PCA,Eigenvalue,Recognition,Biometricsystem,Database,Extraction,Security.*

1. Introduction

The information age is quickly revolutionizing the way transactions are completed. Everyday actions are increasingly being handled electronically, instead of with pencil and paper or face to face. This growth in electronic transactions has resulted in a greater demand for fast and accurate user identification and authentication. When credit and ATM cards are lost or stolen, an unauthorized user can often come up with the correct personal codes. Despite warning, many people continue to choose easily guessed PIN's and passwords of their birthdays, phone numbers and social security numbers. Recent cases of identity theft have heighten the need for security methods

Face detection technology may solve this problem since a face is undeniably connected to its owner expect in the case of identical twins. It's nontransferable. The system can then compare scans to records stored in a central or local database or even on a smart card.

Next technology to be used from the security point of view is the Speech Recognition which can be defined as the process of converting speech signal to a sequence of words by means of algorithm implemented as a computer program .Speech processing is one of the exciting areas of signal processing. It has potential of being important mode of interaction with computer.

This paper gives detailed idea of technique developed in each stage of face recognition and major technological perspective and appreciation of the fundamental progress of speech recognition.In this, we proposed new technique for person identification using fusion of both face and speech which can substantially improve the rate of recognition as compared to the single biometric identification for security system development.

1.1Face Detection

It is a necessary first-step in face recognition systems with the purpose of localizing and extracting the face region from the background. However, it was not until recently that the face detection problem received considerable attention among researchers.

Face recognition means to identify the human face and gives the important information about that person which are available in our database. Fig.1 gives the idea that how the face detection takes place.

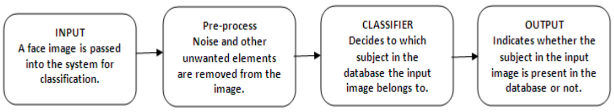


Fig.1 Block diagram Representation of a Face Detection System

1.2 Speech Detection

Speech detection refers to the ability to listen spoken words and identify various sounds present in it and recognize them as words of some known language. It is the process of converting spoken input to text. Speech recognition is thus sometimes referred to as speech-to-text.

*Corresponding author: **Gopal.N.Dhoot**

In order to recognize speech, the system usually consists of pre-processing and post-processing. Pre-processing involves feature extraction and the post-processing stage comprises of building a speech recognition engine. Speech recognition engine usually consists of knowledge about building an acoustic model, dictionary and grammar. Once all these details are given correctly, this engine identifies the most likely match for the given input and it returns the recognized word or utterance. Following fig.2 indicates the representation of it.

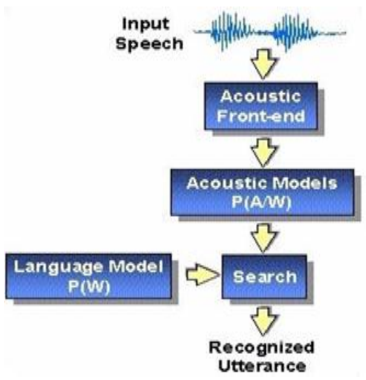


Fig.2 Block Representation of a Face Detection System

The development of personal identification based on face and speech recognition is presented. One problem with face recognition by human is that people find it relatively easier to recognize faces of their own race than other races. But face recognition by machine eliminates the problem of racial subjectivity. People find it very easy to recognize other familiar people by the speech even if they are out of sight. This exploits the machine ability to recognize human face and the possibility of porting man's ability to recognize people by their voices to machine.

The remainder of this paper is organized as follows: Section 2 presents the proposed method. Section 3 introduces two main modules such as face and speech recognition which comes under methodology for detecting these traits. Section 4 describes idea of the multimodal biometric using fusion of both. Next section 5 contains experimental results, graphical analysis and comparison of differentbiometrictraits.Finally,conclusion is given in section 6.

2. Proposed Method

The proposed method of face and speech detection mainly gives the idea about how their recognition takes place. This is basically explained here with the help of two modules such as

i. Face Recognition Module
ii. Speech Recognition Module

Both the modules described here gives detailed idea of their techniques used. In the first module, PCA-DCT method which is used for feature extraction and then verification of faces. Similarly, in the second module, MFCC algorithm is used to extract the features of speech. Finally, we have done fusion of these two traits as face and speech which provides us the combine verified system with great accuracy.

3. Methodology

3.1 Face Recognition Module

3.1.1Facial Image Acquisition (The Database)

The data for this experiment is collected from the publicly available database shown in fig.3 called the ORL Database. There are ten different images of each of 40 distinct subjects or individuals. For some subjects, the images were taken at different times, varying the lighting, facial expressions such as open or closed eyes or smiling or not smiling and having glasses or no glasses. All the images were taken against a dark homogeneous background with the subjects in an upright, frontal position with tolerance for some side movement. The size of each image is 92x112 pixels, with 256 grey levels per pixel.

Fig. 3 Sample images for a subject of the ORL Database

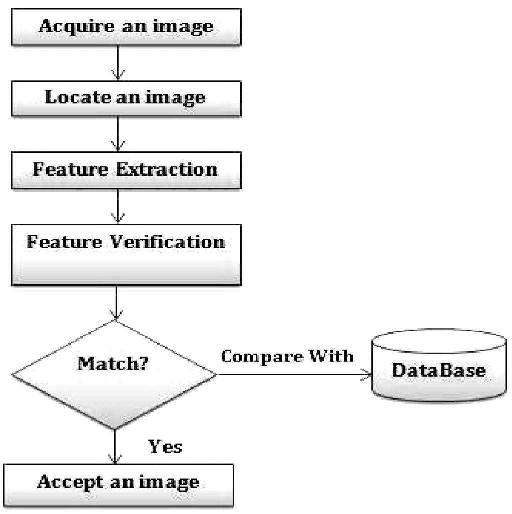


Fig.4 Step by Step approach of Face Recognition System

Fig.4 shows a step by step approach. Principal Component Analysis (PCA), proposed by Turk is one of the most important single sample face recognition methods, which can exactly express every face image via linear operation of eigenvector.

3.2 Face Recognition Problem

During the past decades, face recognition has received substantial attention from researchers. The challenges of it are the rapid and accurate identification or classification of a query image. Rapid can be associated to speed and accuracy refers to recognition rate. Most techniques emphasize on the efficiency in getting positive results but, when it comes to implementation, speed is vital. The performance of a face recognition technique should be able to produce the results within a reasonable time.
eg. For video monitoring and artificial vision, real time face recognition has a very important meaning. It is very useful that the system can detect, recognize and track subject in real time. In human-robot interaction, real-time response time is critical. Besides, it also enables computer systems to recognize facial expressions and infer emotions from them in real time.

3.3 Feature Extraction

Feature extraction is a key step of any face recognition system. It is an important method in the fields of pattern recognition and data mining technology. It extracts the meaningful feature subset from original dates by some rules, to reduce the time of machine training and the complexity of space, in order to achieve the goal of dimensionality reduction. Feature extraction transforms the input data into the set of features while the new reduced representation contains most of the relevant information from the original data. Feature extraction is a process which transfers the data from primary spaces into feature space, representing them in a lower dimensional space with less effective characters. Up to now, many methods of feature extraction have been proposed, such as knowledge-based methods, feature invariant approaches, template matching methods and appearance-based methods. Among them, the algorithm of Eigen face, the most widely used method of linear map based on PCA (Principle Component Analysis) has become the mainstream criterion to test the performance of various face recognition system.

3.4. Principal Component Analysis (PCA)

Principal Component Analysis (PCA) is a dimensionality reduction technique that can be used to solve compression and recognition problems. PCA is also known as Hotelling or eigen space Projection or Karhunen and Leove (KL) transformation. PCA transforms the original data space or image into a subspace set of Principal Components (PCs) such that the first orthogonal dimension of this subspace captures the greatest amount of variance among the images. The last dimension of this subspace captures the least amount of variance among the images, based on the statistical characteristics of the targets. The output components from this transformation are orthogonal or uncorrelated and the mean square error can be the smallest when describing the original vector with these output components.

PCA is a popular transform technique which result is not directly related to a sole feature component of the original sample. It has the potential to perform feature extraction, that able to capture the most variable data components of samples and select a number of important individuals from all the feature components. PCA has been successfully applied on face recognition, image denoising, data compression, data mining, and machine learning. The majority of the applications of PCA are to use PCA to transform samples into a new space and to use lower dimensional representation from the new space to denote the sample. Implementation of the PCA method in face recognition is called eigen faces technique. Turk and Pentland presented the eigen faces method for face recognition in 1991. Face images were projecting onto a face space defined by the eigen faces, and the eigenvectors of the set of faces not necessary corresponded to isolated features such as eyes, ears, and noses etc. The eigen faces algorithm uses PCA for dimensionality reduction in order to find the best account of vectors for the distribution of face images within the entire image space. PCA has been widely investigated. It has become one of the most successful approaches in face recognition and the most fully characterized samples. The procedures of Principal Component Analysis consist of two phases, training step and recognition step.

3.4.1 Training Step

This step is a process to get eigen space from training image which previously has been changed into data matrix. Samples of data, on which the system needs to recognize are used to create an Eigen Matrix which transforms the samples in the image space into the points in eigen space.

3.4.2 Recognition Step

This step is a process to get eigen space from test image which previously has been changed into data matrix. These results were then compared with results from training phase to get minimum difference.

3.5 Eigen Face Approach

3.5.1 Eigen Values and Eigen Vectors

In linear algebra, the eigenvectors of a linear operator are non-zero vectors which, when operated on by the operator result in a scalar multiple of them. The scalar is then called the eigen value (**λ**) associated with the eigen vector (X). Eigen vector is a vector that is scaled by a linear transformation. It is a property of a matrix. When a matrix acts on it, only the vector magnitude is changed not the direction.

$$AX=\lambda X \tag{1}$$

Where, A is a Vector function.

3.5.2 Calculations of Eigen Values and Eigen Vectors

By using (1), we have the equation,

$$(A-\lambda I)\ X=0 \tag{2}$$

Where, I is the n x n Identity matrix.

This is a homogeneous system of equations, and from fundamental linear algebra, we know that a nontrivial solution exists if and only if

$$\det\ (A-\lambda I) = 0 \tag{3}$$

Where, det ( ) denotes determinant.

When evaluated, becomes a polynomial of degree n. This is known as the characteristic equation of A, and the corresponding polynomial is the characteristic polynomial. The characteristic polynomial is of degree n. If A is n x n, then there are n solutions or n roots of the characteristic polynomial. Thus, there are n eigen values of A satisfying the equation,

$$AX_i=\lambda X_i \tag{4}$$

Where i=1, 2, 3….n

If the eigen values are all distinct, there are n associated linearly independent eigen vectors, directions are unique, which span an n dimensional Euclidean space.

In the case where there are r repeated eigen values, then a linearly independent set of n eigenvectors exist, provided the rank of the matrix

$$(A-\lambda I) \tag{5}$$

is rank n-r. Then, the directions of the r eigenvectors associated with the repeated eigen values are not unique.

3.5.3 Face space creation

The accurate reconstruction of the face is not required. So, we can now reduce the dimensionality to M' instead of M. This is done by selecting the M' Eigen faces which have the largest associated Eigen values. These Eigen faces now span a M' dimensional which reduces computational time.

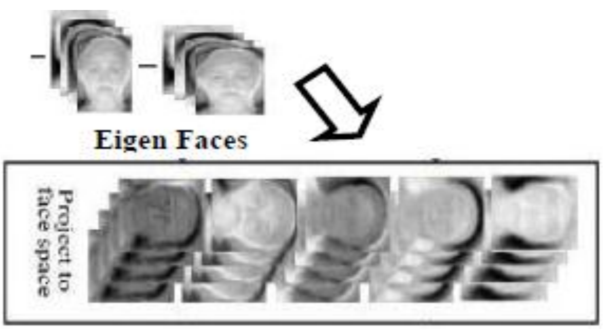


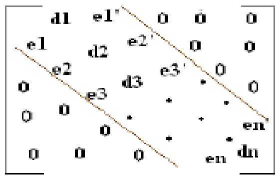
Fig. 5 Face Space

Fig.5 describes about Eigen faces. In order to reconstruct the original image from the eigen faces, one has to build a kind of weighted sum of all eigen faces ie. Face Space. That is, the reconstructed original image is equal to a sum of all eigen faces, with each eigen face having a certain weight. This weight specifies, to what degree the specific feature is present in the original image. If one uses all the eigen faces extracted from original images, one can reconstruct the original images from the eigen faces exactly. But one can also use only a part of the Eigen faces. Then the reconstructed image is an approximation of the original image. However, one can ensure that losses due to omitting some of the eigen faces can be minimized. This happens by choosing only the most important features ie. eigen faces.

3.5.4 Calculation of Eigen Values

Two algorithms called TRED2 () & QL algorithm are used for calculating Eigen Values. In TRED2 algorithm, Covariance Matrix is given as a input. Here Covariance Matrix is converted into Tri diagonalised form except Upper, Lower & main diagonal elements all other elements are made zero.

Consider an example:



3.5.5 Process of Recognition system

3.5.5.1 Eigen faces Initialization

i. Acquire an initial set of face images ie. the training set.
ii. Calculate the Eigen faces from the training set, keeping only the M images that correspond to the highest eigen values. These M images define the face space. As new faces are experienced, the Eigen faces can be updated or recalculated.
iii.Calculate the corresponding distribution in M dimensional weight space for each known individual, by projecting their face images onto the Face Space.

3.5.5.2 Eigen faces Recognition

i. Calculate a set of weights based on the input image and the M Eigen faces by projecting the input image onto each of the Eigen faces.
ii. Determine if the image is a face at all by checking to see if the image is sufficiently close to Face Space.
iii. Update the Eigen faces and/or weight patterns.
iv. If it is a face, classify the weight pattern as either a known person or as unknown.

3.6 Speech Recognition Module

The first task is to identify the presence of a speech signal. This task is easy if the signal is clear, however frequently the signal contains background noise, resulting from a noisy microphone, a fan running in the room, etc. The signals obtained were in fact found to contain some noise. We have used two criteria to identify the presence of a spoken word. First, the total energy is measured, and second the number of zero crossings are counted. Both of these were found to be necessary, as voiced sounds tend to have a high volume and thus a high total energy but a low overall frequency ie. a low number of zero crossings, while unvoiced sounds were found to have a high frequency, but a low volume. Only background noise was found to have both low energy and low frequency. The method was found to successfully detect the beginning and end of the several words tested.

But, it is not sufficient for the general case, as fluent speech tends to have pauses, even in the middle of words e.g. in the word 'acquire', between the 'c' and 'q'. In fact reliable speech detection is a difficult problem and is the important part of speech recognition; however the method that we described below is sufficient for this paper. The speech input is recorded at a sampling rate of 22050 Hz.

This sampling frequency is chosen to minimize the effects of aliasing in the analog-to-digital conversion process. In this work, the Mel frequency Cepstrum Coefficient (MFCC) feature has been used for designing a text dependent speaker identification system.

The Speaker recognition is a generic term used for two related problems: Speaker identification and verification. In the identification task the goal is to recognize the unknown speaker from a set of N known speakers. In verification, an identity claim e.g. a username is given to the recognizer and the goal is to accept or reject the given identity claim. In this work, we concentrate on the identification task. The input of a speaker identification system is a sampled speech data, and the output is the index of the identified speaker.

Steps for speech recognition are as

a.Voice input(i)DataSets(known)(ii) Run Time (unknown)
b.Convert voice into .Wav Form
c.Window the signal
d.Apply Fast Fourier Transform (FFT)
e.Take the magnitude
f.Take logarithm of magnitude
g.Wrap the frequencies according to the Mel scale
h.Take the inverse FFT.

There are three important components in a speaker recognition system: the feature extraction component, the speaker models and the matching algorithm.

3.6.1Mel Frequency Spectral Coefficients (MFCC)

MFCC is based on the human peripheral auditory system. Block diagram of it is as shown in fig.6

The human perception of the frequency contents of sounds for speech signals does not follow a linear scale. Thus for each tone with an actual frequency 't' measured in Hz, a subjective pitch is measured on a scale called the Mel Scale. The mel frequency scale is a linear frequency spacing below 1000 Hz and logarithmic spacing above 1kHz.As a reference point, the pitch of a 1 kHz tone, 40 dB above the perceptual hearing threshold, is defined as 1000 Mel.

The Mel-frequency cepstral coefficients are frequently used as a speech parameterization in speech recognizers.

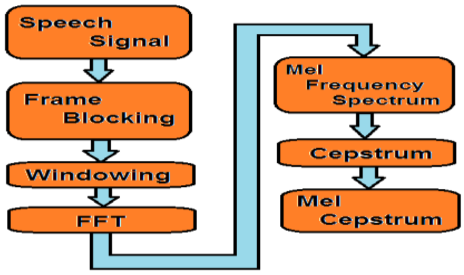


Fig.6 Block diagram of Feature Extraction Process by applying MFCC

Practical applications of speech recognition and dialogue systems bring sometimes a requirement to synthesize or reconstruct the speech from the saved or transmitted MFCCs.

The extracted speech features of a speaker are quantized to a number of centroid using vector quantization algorithm. These centroids constitute the codebook of that speaker. MFCC"s are calculated in training phase and again in testing phase. Speakers uttered same words once in a training session and once in a testing session later. The Euclidean distance between the MFCC"s of each speaker in training phase to the centroid of individual speaker in testing phase is measured and the speaker is identified according to the minimum Euclidean distance.

4. Fusion for Face and Speech

4.1Fusion by the Majority Vote

If the majority of the systems decided 1 then the final decision is YES. Majority Vote is a simple method to combine the exits of multiple sources and use a voting process. In this case, each source must provide a decision of its choice and the final decision is based on a majority rule.

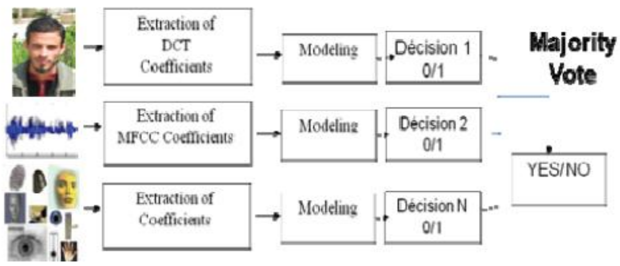


Fig.7 Fusion by the majority vote

5. Performance Analysis

5.1 Experimental Results

The reliability of the proposed multimodal biometric system is described with the help of experimental results. The system has been tested on a database of 400 individuals. The training database contains a face and speech for each individual. The face image has been taken under controlled environment using a digital camera.

The GUI is created by using MATLAB software for face and speech. Fig.8 (a) below indicates the verified result obtained for face and next fig.8 (b) provides the complete verification of these two traits as face and speech.

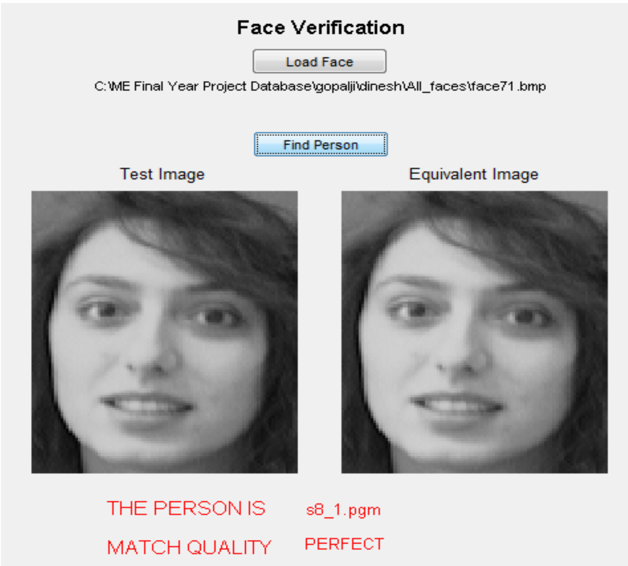


Fig.8 (a) Verified Test image with input Data Set

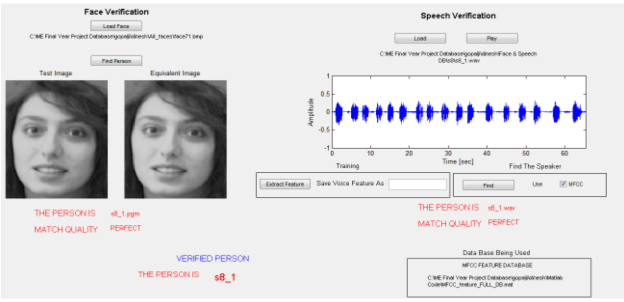


Fig.8 (b) Verified image of Face and Speech

5.2 Graphical Analysis between different Biometric Traits

The graph shows the weighted percentage of biometrics.

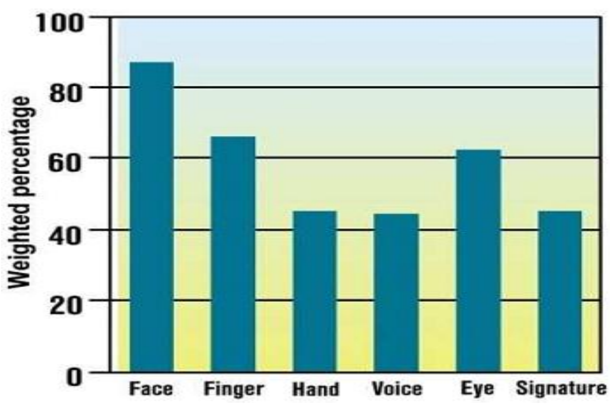


Fig.10 Graphical representation of Biometrics Traits

5.3 Comparison between different Biometric Identification

Table1 Comparison between Biometric Identification

Biometric Patterns	Face	Speech	Iris	Finger prints
Distinctiveness	High	Medium	High	High
Robustness	High	Medium	High	Medium
Accessibility	Medium	High	Low	Medium
Acceptability	High	High	Medium	Medium

Conclusion

Here, we have used PCA method along with Eigen face approach which provides a statistical dimensionality reduction that produces optimal linear least squares decomposition of a training set and MFCC algorithm for Speech Detection process is also used. Therefore it will results in rapid speed which works well under the constrained environment.

From this project, we have heightened the level of security. If any one of face and voice does not match with the database available then the person is denied from getting verification and finally he was not detected.

Person Recognition using these traits as face and voice recognition is still a very challenging topic after decades of exploration. A number of typical algorithms are presented separately, being categorized into appearance based schemes.

Acknowledgement

In this paper, work done by many authors has reviewed. We sincerely like to thank all the people who directly and indirectly motivated and helped us in developing this project as nice one. Also, we extend our thanks to the experts who have contributed in our work.

References

Nick Pears Thomas Heseltine and Jim Austin. Evaluation of image preprocessing techniques for eigenface based face recognition. ACA Group, Dept. of Computer Science, University of York, 2002.

Wikipedia http://en.wikipedia. org/wiki/ Eigenvalue,eigen vector and eigen space

Turk M, Pentland A. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 1991.3(1): 71-86

K.E. Gates, Fast and Accurate Face Recognition Using Support Vector Machines, *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2005, pp.163-163.

S. Palanivel, B.S. Venkatesh, and B. Yegnanarayana, Real Time Face Recognition System Using Autoassociative Neural Network Models, 2003 *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2003. Proceedings.(ICASSP '03), 2003, pp. II-833-6.

S.jen Lin, C.-yang Lee, M.hsuan Chao, C.senChiou, and C.-sing Yang, The Study and Implementation of Real-Time Face Recognition and Tracking System, *Proceedings of the Ninth International Conference on Machine Learning and Cybernetics*, 2010, pp. 11-14.

C. Cruz, L.E. Sucar, and E.F. Morales, Real-Time Face Recognition for Human-Robot Interaction, 2008 *8th IEEE*

International Conference on Automatic Face & Gesture Recognition, Sep. 2008,pp. 1-6.

P. Michel and R. El Kaliouby, Real Time Facial Expression Recognition in Video Using Support Vector Machines, *Proceedings of the 5th international conference on Multimodal interfaces - ICMI'03*, 2003, p. 258.

L. Xie and J. Li, A Novel Feature Extraction Method Assembled with PCA and ICA for Network Intrusion Detection, *2009 International Forum on Computer Science-Technology and Applications*, vol. 3, 2009, pp. 31-34.

M. Karg, R. Jenke, W. Seiberl, K. K, A. Schwirtz, and M. Buss, A Comparison of PCA , KPCA and LDA for Feature Extraction to Recognize Affect in Gait Kinematics, *3rd International Conference on Affective Computing and Intelligent Interaction and Workshops*, 2009, pp. 1-6.

Ö. Toygar and A. Acan, Face Recognition Using PCA , LDA and ICA Approaches on Colored Images, *Journal of Electrical & Electronic Engineering*, vol. 3, 2003, pp. 735-743.

F. Song, Z. Guo, and D. Mei, Feature Selection Using Principal Component Analysis, *2010 International Conference on System Science, Engineering Design and Manufacturing Informatization*, vol. 1, 2010, pp. 27-30.

B. Poon, M.A. Amin, and H. Yan, PCA Based Face Recognition and Testing Criteria, *Proceedings of the Eighth International Conference on Machine Learning and Cybernetics,* Jul. 2009, pp. 2945-2949.

Cepstral Features and Text-Dependent Speaker Identification –A Comparative Study Atanas Ouzounov C& Vol 10. No 1 2010.

Speech Recognition by Machine: A Review, M.A.Anusuya, *International Journal of Computer Science and Information Security*, Vol. 6, No. 3, 2009.

Wavelet Formants Speaker Identification Based System via Neural Network, K. Daqrouq, *International Journal of Recent Trends in Engineering*, Vol 2, No. 5, November 2009.