Volume 4, Issue 9, September 2014 ISSN: 2277 128X

International Journal of Advanced Research in Computer Science and Software Engineering

Research Paper

Available online at: www.ijarcsse.com

Multimodal Behavioural Biometric Personal Authentication Based on Handwritten Signature and Speech

Binsu C. Kovoor*
Department of Computer Science
Cochin University of Science and Technology, Cochin, India

Supriya M. H.
Department of Electronics,
Cochin University of Science and Technology, Cochin, India

K. Poulose Jacob
Department of Computer Science
Cochin University of Science and Technology, Cochin, India

Abstract— In this era of digital impersonation, physiological or behavioural biometric techniques has gained significant interest in the wake of heightened concerns about security and rapid advancements in networking, communication and mobility. The unimodal behavioural biometric techniques are not suited to all users and scenarios. In this study, a multimodal behavioural biometric system utilising speech and handwritten signature has been implemented and error analysis has been carried out. A total of one thousand speech data and signature images of 100 users are used for training the proposed system. The short term spectral features are extracted from the sound data and Vector Quantization was done using K-means algorithm. The static and dynamic features of online handwritten signature identification algorithm are obtained. The static features of a signature image are extracted by grid based Gabor wavelet. The dynamic features are derived from values in the output stream produced by the signature tablet. A multimodal user dependent weighted fusion algorithm combines the results obtained from the speech data and signature data. The results showed that accuracy of multimodal system using speech and signature is higher when compared to individual unimodal recognition and improve the classification performance with an overall EER of 6 percent.

Keywords— Multimodal Biometrics, Handwritten Signature, Speech, Gabor wavelet, Spectral, Cepstral.

I. INTRODUCTION

The establishment of the identity of a person in a reliable and time-efficient manner is of paramount importance as the society is becoming increasingly dependent on the use of information technology for everyday tasks [1]. Of the many automatic identification technologies, the methods based on biometrics have gained considerable attention due to their robustness as well as reliability. Biometrics is a measurable distinctive physical characteristic or personal trait that can be used to identify an individual or to verify the claimed identity of an individual [2]. An optimal biometric system is one having the properties of distinctiveness, universality, permanence, acceptability, collectability, and security [3]. However there is no single biometric identifier which has all of these properties. As a solution, multiple biometric identifiers are incoporated in a single system, commonly referred to as multimodal biometric system. Multibiometric systems integrate evidence from multiple sources of biometric information inorder to authenticate the identity of an individual [4]. The system reliability increases when multiple traits are being accounted for in the identification process [2]. The limitations of unibiometric systems can be alleviated by multibiometric systems. Multibiometric systems reduce the effect of noisy data thereby enabling reliable determination of identity even if one of the biometric samples is noisy. Multibiometric systems are resistant to spoof attacks since it is difficult to simultaneously spoof multiple biometric sources. Further, a multibiometric can also check the liveness of the users by acquiring a subset of traits in some random order.

The integration of different biometric sources, often termed as biometric fusion, is another main design issue. It has a good impact on the performance of the system. The fusion scheme can be classified into sensor level, feature level, score level and decision level. The choice of fusion depends on the type of information from the biometric sources namely, raw biometric samples, feature sets, match score and decision labels. In this paper two behavioral biometric traits, signature and voice are integrated for identification.

Voice of a person holds certain unique characteristics which can be utilized for personal authentication. Identifying the exact person who is speaking is the most popular biometric security system. Voice is a very intuitive behavioral and ubiquitous biometrics which can be captured by modern personal computer and requires no expensive special hardware other than a microphone. Speaker recognition uses the acoustic features of speech that have been found to differ among individuals. These acoustic patterns reflect both anatomy (e.g., size and shape of the throat and mouth) and learned behavioral patterns (e.g., voice pitch, speaking style)[5]. This incorporation of learned patterns into the voice templates has earned speaker recognition and its classification as a behavioral biometric.

Handwritten signature authentication is the process of verifying the identity of a person based on user's handwritten signature [6][7]. Signature has been widely accepted as a means of legal and commercial transactions identity authentication [8]. Signatures have played a historical role in authenticating documents. Being part of everyday life, signature based authentication is remarked as a consistent non-invasive authentication procedure by the majority of the users, therefore, it can help in overcoming some of the privacy difficulties. The main drawback of biometrics when

compared with conventional methods is that many biometrics can be copied or forged [3][8]. Whereas it is always possible to obtain a new key or another password, it is not possible to replace any biometric data [9]. Nevertheless, signature is considered an exception where users can be asked to change their signature if needed. Based on the method used to capture the signatures, handwritten signature biometrics system is divided into two categories, namely, offline and online [10]. The analysis of features extracted from scanned images of handwritten signatures is referred to as off-line or static. The analysis of handwritten signatures captured via digitizing tablets or other electronic devices, which captures the trajectory, pressure and velocity of handwriting is referred to as on-line or dynamic. The main task of any signature verification task is to detect whether the signature is genuine or forged [7]. The two main categories of forgeries are casual or random forgeries and skilled or traced forgeries Casual or random forgeries are attempts of recreating signature trajectory without prior knowledge about the signature style. The skilled forgery is a suitable imitation of the genuine signature. The skilled forgeries are more difficult to detect than random forgeries as the characteristic features of a skilled and traced forgeries resemble closely those of the original signature.

This paper proposes an efficient multimodal behavioral biometric system. In order to increase the performance of individual biometric trait, the individual traits are fused at matching score level using user dependent weighted fusion method.

II. THEORY AND BASIC DEFINITIONS

The multimodal behavioral biometric system is developed using two traits, voice and signature as shown in Fig. 1. The process of user identification can be divided into two main phases, namely, the training phase or enrolment phase and the testing phase or identification phase. During the user enrolment phase, speech and signature samples that contain the discriminating features are collected from the user and feature vectors are generated which are used to train the model. In the recognition phase, the feature vectors extracted from the unknown person's speech and signature are compared against the model in the system database to find the similarity score, for the purpose of making decision. Feature selection is of great importance in multibiometric recognition, as accuracy is highly dependent on the type and number of features used.

In Voice Recognition, the speaker is recognized by combining spectral features and Mel Frequency Cepstral Coefficient (MFCC). In Signature Verification, feature vector consisting of static and dynamic features of signature image is extracted and matched using Euclidean Distance measure. The modules based on the individual traits returns an integer matching score value after matching the database and query feature vectors. The final score is generated by using user dependent weighted fusion technique at matching score level which is then passed to the decision module. The brief description of various recognition algorithms are presented below.

A. Speaker Recognition

Speaker recognition is considered as recognizing person from their voices. The sounds of two individuals sounds are not identical because of the physical differences such as their vocal tract shapes, larynx sizes etc. and also due to the characteristic manner of speaking like the accent, rhythm, pronunciation pattern etc. [5]. Features have been computed from the spectrum of the speech signal and relates directly to some perceptual characteristics of sound, such as loudness, pitch, etc. Most of the features are generated from the spectrogram on a frame-by-frame basis.

1) Feature Extraction The short-term spectral feature, as the name suggests, is computed from short frames of about 20-30 millisecond duration of a speech signal changing continuosly due to articulatory movements. Within this interval, the signal is assumed to remain stationary and the spectral feature vector is extracted for each frame. Typical standard features considered in this prototype for speaker recognition includes Spectral Centroid, Spectral Roll off, Spectral Flux and MFCCs.

Spectral Centroid: The spectral centroid [11], is the centroid of the magnitude spectrum of short time fourier transform and is a measure of spectral brightness. This simple, yet efficient parameter is estimated by summing together the product of each frequency component of the spectrum and its magnitude and then normalized by dividing the result by the sum of all the spectral magnitudes. Thus, the spectral centroid SC is given by Eq.(1) where S_k is the magnitude spectrum of the k^{th} frequency component f_k and N is the record size.

$$SC = \frac{\sum_{k=0}^{N/2-1} f_k S_k}{\sum_{k=0}^{N/2-1} S_k} \qquad (1)$$

Spectral Roll off: Another spectral feature, which gives a measure of the spectral shape, is the spectral roll off [11] and is defined as the frequency below which 85% of the magnitude distribution of the signal is concentrated

$$i.e.\ \ RO = Minimum(R),\ \text{such that}$$

$$\sum_{k=0}^{R} S_k \geq 0.85 \sum_{k=0}^{N-1} S_k \qquad (2)$$

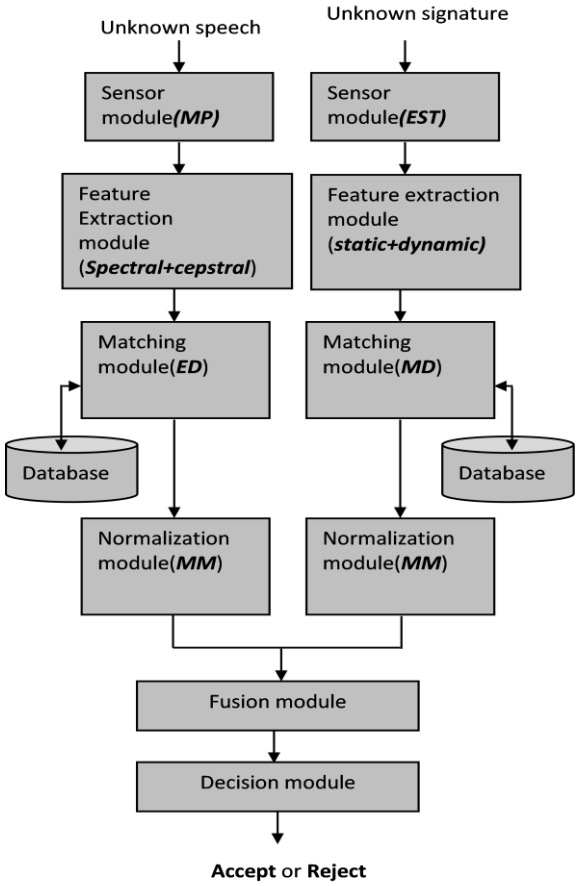

Fig 1. Conceptual block diagram of proposed bimodal system

Spectral Flux: Spectral flux [11], a measure of the amount of local spectral change, can be defined as the squared difference between the normalized magnitude spectra of successive frames as in Eq.(3). In Eq.(3), $norm_f$ is the magnitude spectrum of the current frame, scaled to the range 0 to 1 and $norm_{f-1}$ is the normalised magnitude spectrum of the previous frame. Spectral flux is a measure of how quickly the power spectrum of a signal is changing and is computed by comparing the power spectrum of one frame with that of the previous frame.

$$Flux = \sum (norm_f[i] - norm_{f-1}[i])^2 \qquad (3)$$

Mel Frequency Cepstral Coefficient: The Mel Frequency Cepstral Coefficients (MFCC) [11] are computed with the aid of a psychoacoustically motivated filterbank, followed by logarithmic compression and discrete cosine transform(DCT). The outputs of M channel filterbank are denoted as Y(m), where m=1…M. The MFCCs can be computed from Eq.(4). In Eq.(4), n is index of the cepstral coefficient

$$C_n = \sum_{m=1}^{M} [\log Y(m)] \cos\left[\frac{\pi n}{M}\left(m - \frac{1}{2}\right)\right] \qquad (4)$$

2) Speaker Modelling. Vector Quantization (VQ) model also known as centroid model, is one of the simplest text-independent speaker models [12]. VQ maps the large set of extracted short term spectral feature vectors in to a finite number of clusters, each represented by its centroid. The clustering is done by K-means clustering algorithm and the resulting reduced set of feature vectors is known as codebook. A speaker database is developed consisting of N codebooks, one for each speaker [12]. In recognition phase, the test utterance features denoted as $X=\{x_1 \ldots x_T\}$ of unknown speaker are compared with all the reference vectors denoted as $R=\{r_1 \ldots r_k\}$ of known speakers in the database. The average quantization distortion is given in [10] and defined as in Eq.(5). In Eq.(5), d(.,.) is the Euclidean distance and a smaller value of Eq.(5) indicates higher likelihood for X and R originating from the same speaker.

$$D_q(X, R) = \frac{1}{T} \sum_{t=1}^{T} \min_{1 \le k \le K} d(x_t, r_k) \qquad (5)$$

B. Signature Recognition

The image of signature trajectory is captured using electronic signature tablet. The pre-processing module is responsible for the preconditioning of signature image. The distinct features of the signature image are extracted in feature extraction module. The images are to be pre-processed to make them fit for processing. Initially, the image is binarized and dilation is applied to fill the gaps and broken necks. The image is then thinned and edges are pruned.

1) Feature Extraction: The discriminative power of the features in the reference set plays a major role in the entire identification process. It is important to find the features that are invariant with respect to slight changes in intra-class signatures. The features should be powerful enough to discriminate other signatures in the knowledgebase. In this paper, both static and dynamic features of an online signature are extracted. The texture and topological features are the static features of a signature image. The dynamic features are the features captured by the digital tablet in real-time such as the velocity values, breakpoints, and the time taken to create a signature [14]. The static feature from signature images is extracted using Gabor Wavelet Transform (GWT).

Static Feature: The topological features from the signature image is derived using Gabor wavelet filter [15] at point (x,y) is defined as in Eq.(6)

$$g(x, y, \lambda, \theta, \varphi, \sigma, \gamma) = \exp\left(\frac{x'^2 + \gamma^2 y^2}{2\sigma^2}\right)\cos\left(\frac{2\pi x'}{\lambda} + \varphi\right) \quad (6)$$

In the two dimensional GWT, x' and y' are given in Eq.(7).

$$x' = x\cos\theta + y\sin\theta$$
$$y' = -x\sin\theta + y\cos\theta \quad (7)$$

θ is the orientation of the normal to the parallel stripes of a Gabor function. λ is the wavelength of the cosine wave. φ is the phase offset in the argument of the cosine factor of the Gabor function. σ is the radius of the Gaussian. γ is the spatial aspect ratio of the Gaussian [15].

The feature is extracted from the signature image by placing a virtual grid on signature image. The grid size has to be chosen very carefully. It can neither be too gross nor be too detailed. The Gabor coefficients are computed on each point of grid by convolution. Convolution between Gabor filter and a sub image around point (x,y) is calculated. In each point of virtual grid, 6 complex Gabor coefficients are computed corresponding to $\lambda \in \{2, 2\sqrt{2}\}$ and $\theta \in \{0, \pi/4, \pi/2\}$. Other Gabor filter parameters are assumed to take values of $\varphi \in \{0, \pi/2\}$, $\sigma = 2\lambda$ and $\gamma = 0.5$. This means that for each grid point, two frequencies in three orientations and two phases are computed. Therefore, for all grid points of an image, Gabor coefficients are computed. The absolute values of Gabor coefficients constitute the static feature vector of the signature image.

Dynamic Features: The static image of the signature on a paper can be forged easily. The forgers can reproduce the image (or shape) of a signature, but it is difficult to forge the motions that caused the image [14]. When a signature is captured with a signature tablet, the pen motions which are dynamic in nature are recorded. The values in the output stream produced by the signature tablet are equidistant in time. It contains the x and y coordinates sampled at timestamp t and is represented as x(t) and y(t), respectively. At each sample point, the signature data as S(t) = [x(t), y(t), timestamp(t)], t = 1,…,N , where N is total the number of samples of the signature trajectory along with the timestamp. A sample signature, its x and y plot and its normalized x plot is given in Fig. 2. Average velocity in the X plane (S_{vx}) and Y plane (S_{vy}) are given in Eq.(8) and Eq.(9).

$$S_{vx} = \frac{1}{N}\sum_{i=1}^{N-1}((x_{i+1} - x_i)/(t_{i+1} - t_i)) \quad (8)$$

$$S_{vy} = \frac{1}{N}\sum_{i=1}^{N-1}((y_{i+1} - y_i)/(t_{i+1} - t_i)) \quad (9)$$

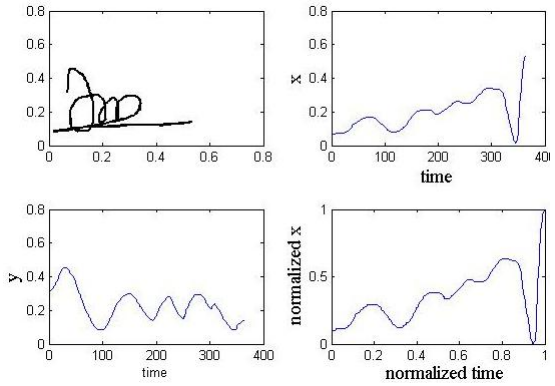

Fig 2. x and y plot of a sample signature.

2) Matching Module. Mahalanobis Distance (MD) computed based on correlation between two signatures is used to verify the similarity of images. When the Mahalanobis distance between the feature vectors of enrolled and tested signature is smaller, then the similarity between the compared signature is higher. The signature raw matching score S_{sg} is the MD between two signatures. S_{sg} is computed as in Eq.(10). In Eq.(10) x and y denote the enrolled and test feature vector respectively. S is the covariance matrix.

$$S_{sg}(\bar{x},\bar{y}) = \sqrt{(\bar{x}-\bar{y}) \quad S^{-1} \quad (\bar{x}-\bar{y})} \tag{10}$$

C. Multimodal Fusion

The fusion technique employed in this work is on the basis of the different weights assigned to each biometric trait. These different weights are computed based on the Equal Error Ratio (EER). The weight W_i for the i^{th} particular trait is calculated using the Eq.(11).

$$W_i = \frac{1/EER_i}{\sum_{j=1}^{n}\left(1/EER_i\right)} \tag{11}$$

where EER_i, the Equal Error Ratio for j^{th} trait and n represents the number of traits participating in fusion. The fused score is calculated as in Eq.(12).

$$S = \sum_{j=1}^{n} W_j \quad x \quad S_j \tag{12}$$

where S_j is the match score of j^{th} trait

III. METHODOLOGY

In the proposed system, the main goal is to evaluate the performance of the behavioural multimodal biometric system based on user dependent weighted fusion approach. In the enrolment phase, speech data and signature are acquired first and then processed according to the training and classification algorithms. In speaker recognition, feature vector of the speech data is derived from spectral and MFCC coefficients. The feature vector is the voice template in the knowledgebase. In signature recognition the feature vector is combination of static and dynamic features. The static features is generated using 2D Gabor filter and dynamic features such as x and y stroke, average velocity in x and y directions. The feature vector thus obtained is the signature template in the knowledgebase. In the identification phase, the matching score of the test template and the training templates are derived. The matching score of speech data are calculated measuring the Euclidean distance between the test and templates in the database. Mahalanobis distance is used for calculating the matching score of signature image. Non-normalized scores cannot be integrated in their raw form, as it is impossible to fuse incomparable numerical scales. The min-max normalization technique is employed in this work to normalize the matching score. These normalized scores are fused using user dependent weighted fusion method.

The proposed algorithm was tested using the speech and signature database as shown in Table I. The database consists of 1200 signature images and speech data of 100 different individuals. Out of the collected data 1000 samples from both signature and speech are used for training the system and rest 200 are used for testing. The system is also tested with 200 samples of unregistered data.

TABLE I: DETAILS OF NUMBER OF USERS

Type of users	**Number of samples**	
	Training	**Testing**
Registered (100 persons)	*100x 10 = 1000*	*100x2=200*
Unregistered		*100x2=200*
TOTAL	1000	400

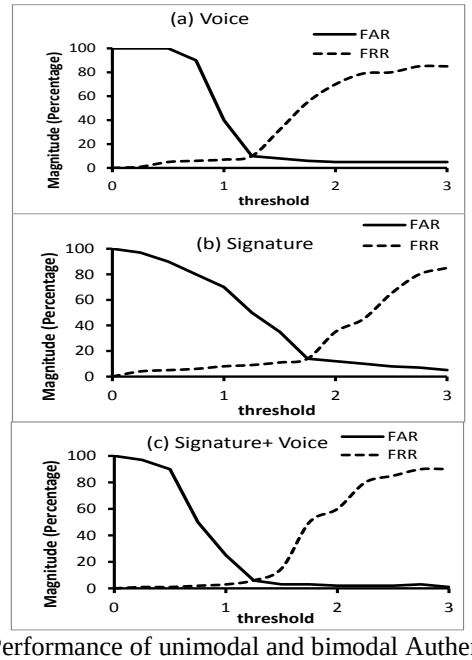

Fig 3. Performance of unimodal and bimodal Authentication

IV. RESULTS AND DISCUSSIONS

The Multimodal behavioural biometric security system based on speech and signature data is implemented in MATLAB. The performance of a biometric system requires some parameters. A decision made by a biometric system is either a "genuine individual" type of decision or an "impostor" type of decision. The two common error rates are False accept rate (FAR) and False Reject Rate (FRR). FAR is defined as "the probability of an impostor being accepted as a genuine individual [16]. That is, in a biometric authentication system, the FAR is computed as the rate of number of people falsely accepted over the total number of enrolled people for a predefined threshold. FRR is defined as "the probability of a genuine individual being rejected as an impostor" [9]. That is, in a biometric authentication system, the FRR is computed as the rate of number of people falsely rejected (genuine people are rejected) over the total number of enrolled people for a predefined threshold. FAR and FRR can be changed by a significant amount depending on the threshold used in the system. If a lower threshold is used in a similarity based biometric matching system, then the FAR will be higher and the FRR will be lower and vice versa. The performance of a biometric system may also be expressed using Equal Error Rate (EER) which refers to that point in a FAR-FRR plot where the FAR equals the FRR. A lower EER value thus indicates better performance [16].

The influence of spectral and cepstral features extracted from the voice signals for speaker identification has been studied. The EER, FAR and FRR has been computed and compared for evaluating the system and is shown in Fig. 3(a). The EER of the system is 10 percent. Hence the accuracy of the unimodal system using speech is 90 percent.

In signature recognition, both static and dynamic features are used for generating feature vector. In this study, the grid size 32x64 has chosen for calculating the Gabor coefficients for feature vector. This feature vector is concatenated with the dynamic features namely, average velocity and normalized x and y stroke area. It has been found from the FAR-FRR graph of Fig. 3(b) that the unimodal signature recognition the EER is 14 percent and hence accuracy is 86 percent. In bimodal system, where the signature and speech scores are combined using weighted fusion method the EER is found to reduce to 6 percent which is shown in Fig. 3(c). Hence the accuracy of the identification system is increased.

V. CONCLUSIONS

Multimodal Biometrics systems are widely used to overcome the traditional methods of authentication. The unimodal behavioural biometric system fails in case of noisy biometric data for particular trait. Thus the user dependent weighted individual scores of two behavioural biometric traits, namely, speech and signature are combined at match score level to develop a multimodal biometric system. The accuracy curve shows that multimodal system performs better as compared to unimodal biometrics with accuracy of more than 94%.

REFERENCES

[1] Binsu C. Kovoor, M. H. Supriya and K. Poulose Jacob, "A Prototype for a Multimodal Biometric Security System Based on Face and Audio Signatures", *International Journal of Computer Science(IJCSc)*, vol 2, No.1, pp 143-147, Jan-June 2011.

[2] A. K. Jain, A. Ross and S. Prabhakar, "An introduction to biometric recognition." *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 1, pp. 4-20, 2004.

[3] A. K. Jain, A. Ross, S. Pankanti, "Biometrics: A tool for information security," *IEEE Transaction on Information Forensics and Security*, Vol.1(2), pp.125-143, 2006

[4] A. Ross and A. K. Jain, "Multimodal biometrics: An overview," *Proc. of 12th European Signal Processing Conference (EUSIPCO)*, pp. 1221-1224, Sep 2004.

[5] R. Lawrence and J. Bing-Hwang, *Fundamentals of Speech Recognition*, New Jersey: Prentice Hall, 1993.

[6] F. Leclerc, R. Plamondon, "Signature verification: The state of the Art," *International Journal of Pattern Recognition and Artificial Intelligence*, Vol. 8(3), pp. 643- 660, 1994

[7] R. Plamondon, and G. Lorette, "Automatic signature verification and writer identification- the state of the art," *Pattern Recognition*, vol. 22, pp. 107-131,1989.

[8] T. Hastie, E. Kishon, M. Clark, and J. Fan, "A Model for Signature Verification*," Proc. IEEE Conf. Systems, Man, and Cybernetics*, pp. 191-196, 1991.

[9] A. Jain and U. Uludag, "Hiding fingerprint minutiae in images," *Proceedings of Third Workshop on Automatic Identification Advanced Technologies*, pp. 97-102, 2002.

[10] R. Plamondon and S. N. Srihari, "On-line and Offline Handwriting Recognition: A Comprehensive Survey," *IEEE Trans. Pattern Recognition and Machine Intelligence*, vol 22(1), 63-84, January 2000.

[11] M.H. Supriya, K.Shaheer, M.G. Mahendran, and P.R.S. Pillai, "Towards Improving the Target Recognition Using a Hierarchical Target Trimming Approach*," WSEAS Transactions on Signal Processing*, Vol.3, pp. 340-345, 2007.

[12] T. Kinnunen, and H. Li, "An overview of text independent speaker recognition from features to supervectors," *J Speech Comm*, vol. 52, pp.12-40, 2009.

[13] M. Shahneh, and A. Taheri, "Voice Command Recognition System Based on MFCC and VQ Algorithms," *Wor Acad of Sc, Eng and Tech*, vol. 57, pp.534-538, 2009.

[14] Y. Sato and K. Kogure, "On-Line Signature Verification Based on Shape, Motion and Writing Pressure," *Proc. Sixth Int'l Conf. Pattern Recognition*, pp. 823-826, 1982.

[15] W. Jiang, K. Lam and T. Z. Shen, "Edge Detection using simplified Gabor wavelet," *IEEE Int. conference Neural Networks and Signal Processing*, pp. 586-591, 2008

[16] A. Ross, K. Nandakumar, and A. K. Jain, *Handbook of multibiometrics*. New York: Springer-Verlag, 2006.

