Multimodal Biometric Person Authentication using Speech, Signature and Handwriting Features

Eshwarappa M.N.
Telecommunication Engineering Department
Sri Siddhartha Institute of Technology
Tumkur-572105, Karnataka, India

Dr. Mrityunjaya V. Latte
Principal and Professor
JSS Academy of Engineering and Technology
Bangalore-560060, Karnataka, India

***Abstract*—The objective of this work is to develop a multimodal biometric system using speech, signature and handwriting information. Unimodal biometric person authentication systems are initially developed for each of these biometric features. Methods are then explored for integrating them to obtain multimodal system. Apart from implementing state-of-the art systems, the major part of the work is on the new explorations at each level with the objective of improving performance and robustness. The latest research indicates multimodal person authentication system is more effective and more challenging. This work demonstrates that the fusion of multiple biometrics helps to minimize the system error rates. As a result, the identification performance is 100% and verification performances, False Acceptance Rate (FAR) is 0%, and False Rejection Rate (FRR) is 0%.**

Keywords- Biometrics; Speaker recognition; Signature recognition; Handwriting recognition; Multimodal system.

I. INTRODUCTION

In the present era of e-commerce more and more services are being offered over the electronic devices and internet. These include banking, credit card facility, e-shopping, etc. To ensure proper use of these facilities only by the authorized or genuine users and avoid any misuse by the unauthorized or imposter users, some person authentication scheme is embedded into these services. Currently, person authentication is done mostly using one or more of the following means: text passwords, personal identification numbers, barcodes and identity cards. The merit of these schemes is that they do not change their value with respect to time and also unaffected by the environment in which they are used. The main demerit of them is that they can be easily misused or forgotten. Also, with time more and more services are being offered over the electronic devices and internet. Hence it becomes unmanageable to keep track of the authentication secrets for different services. The alternative that provides relief from all these demerits is the use of biometric features for person authentication. Any physiological and/or behavioural characteristics of human can be used as biometric feature provided it possesses the following properties: universality, distinctiveness, permanence, collectability, circumvention, acceptability and performance [2].

Some of the commonly used biometric features include speech, face, signature, finger print, handwriting, iris, DNA, Gait, etc. In practice, no single biometric can satisfy all the desirable characteristics mentioned above for it to be used for person authentication. This is due to the problems associated with noisy data, intra-class variation, non-universality, spoof attacks and high error rates [2]. To overcome this limitation, multiple biometric features can be used for person authentication. This resulted in the development of multimodal biometric person authentication system [2]. Thus biometric system can be classified as unimodal system and multimodal system based on whether single or multiple biometric features are used for person authentication. Biometric security system becomes a powerful tool compared to electronics based security systems [1]. Biometrics is fast becoming applicable in various walks of life. Basically, it deals with the use of computer technology and signal processing to identify people based on their unique physical and behavioural characteristics such as fingerprints, voice scans, retinal patterns, facial characters and human DNA mapping. Typically, a biometric system comprises a sensor, interface and a signal processor with driver software. The various different biometric procedures fall into two categories: Static process relating to the identification of fingerprints, hand geometry, Iris or retina and face, and Dynamic processes relating to the recognition of handwriting, keyboard typing patterns, voice, lip movement and behaviour analysis.

A biometric sensor works on the inputs provided by any of the human characteristics and applies an algorithm on the scanned biometric data. This is then compared with, and matched to, a template that has already been created earlier and approved by the user. The most specific and reliable biometric data is obtained from the DNA sequencing of any subject. The matching and comparing process creates a 'score' based on how closely the sampled biometric matches with the template already obtained. A match score is known as genuine score if it is a result of matching two samples of a biometric trait of the same user. It is known as an imposter score if it is the result of matching two samples of a biometric trait originating from different users. An imposter score that exceeds the predefined threshold results in a false accept, while a genuine score that falls below the predefined threshold results in a false reject. The False Accept Rate (FAR) of a biometric system is the fraction of imposter scores exceeding the threshold. Similarly, the False Reject Rate (FRR) of a system is defined as the fraction of genuine scores falling below the threshold. Regulating the value of threshold changes the FRR and the FAR values, but for a given biometric system,

it is not possible to decrease both these errors simultaneously. In real-world biometric system, biometric measure is referred in terms of FAR and FRR. The FAR measures the percentage of invalid users who are incorrectly accepted of genuine users and the FRR measures the percentage of valid users rejected as imposters. The Equal Error Rate (EER) refers to the point where the FAR equals the FRR. Lower the value of EER, the more accurate the biometric system.

There are several multimodal biometric person authentication systems developed in the literature [2-12]. Person authentication based on speech and face features, is one of the first multimodal biometric system [3]. Two acoustic features from speech and three visual features from face are used to build a multimodal system. Multimodal system using face and fingerprint features is then proposed [4]. Finger print verification on top of the face recognition is used to improve the recognition accuracy. The use of clustering algorithms for the fusion of decisions from speech and face modalities are explored [5]. A practical multimodal system using face, voice and lip movement is then developed [6]. The focus is on improving security by considering a dynamic feature like lip movement. In 2004, A. K. Jain et.al., proposed the framework for multimodal biometric person authentication [2]. They discussed in detail about the significance of biometric person authentication and desirable characteristics for a physiological and/or behavioural characteristics of human to be useful as biometric feature. Several biometric features in use are described in terms of these characteristics to highlight the strength and weaknesses of each of them. Details about the different levels of fusion, security and privacy concerns are also discussed. In recent times much of the interest is in audio-visual multi biometric systems [12].

The objective of our work is to develop a multimodal biometric person authentication system using speech, signature and handwriting biometric features. The motivation for the same are explained as follows: (1) speech is both physiological and behavioural, and signature and handwriting are behavioural biometric features, (2) each of these biometric features can be collected using sensors which are cheap and provide reasonably good quality data. All these features are non-intrusive type, easy to collect and hence acceptability among users will be high, (3) there are several practical applications where these three modalities fit in very well, like banking transaction. To complete a financial transaction, you can write the required amount using an electronic pen on the cheque displayed onscreen and put your signature. This enables giving both handwriting and signature features. You can read the amount written and other details that provide speech data, (4) speech is one of the mostly explored biometric features by the speech processing community for the development of speaker recognition system. Speech biometric feature will immensely benefit from the developments available in the speaker recognition literature, (5) the recent trend in human-computer interaction is the electronic-pen based input to the computer that includes Personal Digital Assistant (PDA) and Tablet-Personal Computers (PCs).

With this integrated input device, we have an easy way of capturing signatures and handwriting information. Thus these features can also be integrated into the multimodal system along with speech, (6) most of the existing signature verification systems are online type that uses dynamic features like time and pressure information. However, development of offline signature verification system may benefit from the rich image processing techniques available. A hybrid system using online and offline features may then be developed for increased robustness, (7) most of the handwriting recognition system is meant for forensic investigation. However, handwriting may also have significant information for person authentication. A detailed exploration is required from this perspective, (8) a person authentication system using handwriting information may also benefit from the speaker recognition literature by drawing parallels between the two. It may also be possible to extract some synchronous features between speech and handwriting to reduce spoof attacks.

The present work mainly deals with the implementation of multimodal biometric system employing speech, signature and handwriting as the biometric modalities. This includes feature extraction techniques, modelling techniques and fusion strategy used in biometric system. The organization of the paper is as follows: Section II deals with speaker recognition system, signature recognition system and handwriting recognition system using different feature extraction and modelling techniques, and Section III deals with multimodal biometric person authentication system by combining speaker, signature and handwriting recognition systems using fusion strategy. Section IV provides conclusion and suggestion for future work.

II. Development of Unimodal Systems

A. Speaker Recognition System

Speaker recognition is the task of recognizing speakers using their speech signal. The unimodal biometric system using speech analyzes and extracts speaker-specific features from the speech signal. The extracted features are then separately modeled to obtain one reference model for each speaker. During testing same analysis and feature extraction are carried out to extract speaker-specific features. These features are compared with the reference models to decide on the speaker. The speaker of the reference model that matches closely with the test speech features is declared as the speaker. In person authentication case, claimed identity is given along with the test speech. Hence comparison is done only with the claimed identity reference model and the claim is accepted or rejected based on the comparison with a preset threshold.

The state of the art system builds a unimodal system by analyzing speech in blocks of 10-30 milli seconds with shift of half the block size. Mel Frequency Cepstral Coefficients (MFCCs) are the mostly used features extracted from each of the blocks [13]. The MFCCs from the training or enrollment data are modeled using Vector Quantization (VQ) technique [14]. The MFCCs from the testing or verification data are compared with the VQ to validate the identity claim of the speaker. The MFCCs represent mainly the vocal tract aspect of speaker information and hence take care of only physiological aspect of speech biometric feature. Another important physiological aspect contributing significantly to speaker characteristics is the excitation source [16]. The behavioral biometric aspect of speech is present at longer duration levels

that can be characterized using supra-segmental features like speaking rate, pitch contour, duration etc. In this work, apart from the development of conventional speaker recognition system using MFCC and VQ features termed as baseline system, methods will also be explored to model excitation source and supra-segmental speaker-specific features. These features are then integrated into the baseline system. This may result in an improved and robust speaker recognition system.

Speaker recognition is the task of recognizing the speakers using their voices [17]. Speaker recognition can be either identification or verification depending on whether the goal is to identity the speaker among the group of speaker or verify the identity claim of the speaker. Further, speech from the same text or arbitrary text may be used for recognizing the speakers and accordingly we have text dependent speaker identification and verification approaches. The present work approaches text dependent speaker identification and verification of a speaker through identification. In this work, two different feature extraction and modeling techniques are used for text dependent speaker recognition. The feature extraction techniques are: (1) Mel Frequency Cepstral Coefficients (MFCC) are derived from cepstral analysis of the speech signal, (2) a new feature set, named the Wavelet Octave Coefficients of Residues (WOCOR), is proposed to capture the spectro-temporal source excitation characteristics embedded in the linear predictive residual of speech signal [16]. The two modeling techniques are used for modeling the person information from the extracted features are: (1) Vector Quantization (VQ), (2) Gaussian Mixture Modeling (GMM).

1) Feature extraction phase.

The speaker information is present both in vocal tract and excitation parameters [18]. The MFCCs represent mainly the vocal tract aspect of speaker information and hence take care of only physiological aspect of speech biometric feature. The vocal tract system can be modeled as a time-varying all-pole filter using segmental analysis. The segmental corresponds to processing of speech as short 10 to 30 milliseconds overlapped 5 to 15 milliseconds windows.

The vocal tract system is assumed to be stationary within the window and is modeled as an all-pole filter of order P using linear prediction analysis. The feature vectors that are extracted from smooth spectral representations are cepstral coefficients. In the present work we are using MFCC as feature vectors. The cepstral analysis used for separating the vocal tract parameters and excitation parameters of speech signal s(n). This analysis uses the fundamental property of convolution. The cepstral coefficients (C) are derived by using Fast Fourier Transform (FFT) and Inverse FFT (IFFT) which is given by equation (1).

$$C = real(IFFT(\log|IFFT(s(n))|)) \qquad (1)$$

Human auditory system does not perceive the spectral components in linear scale, but it will perceive on a nonlinear scale. So we can use the nonlinear scale, Mel frequency scale, to extract the spectral information. The critical band filters are used to compute the MFCC feature vectors by mapping the linear spaced frequency spectrum (f_{HZ}) into nonlinearly spaced frequency spectrum (f_{Mel}) using equation (2).

$$f_{Mel} = 2595\log_{10}\left(1+\frac{f_{HZ}}{700}\right) \qquad (2)$$

When a speech signal is given as an input to the feature extractor, it will truncate entire speech signal into frames of length 10-30 ms to make it quasi-stationary. Hamming window is used for eliminating the Gibbs oscillations, which occur by truncating the speech signal. But, due to windowing, samples present at the verge of window are weighted with lower values. In order to compensate this, we will try to overlap the frame by 50%. After windowing, we compute the log magnitude spectrum of each frame and calculating the energy in each critical filter bank. After finding the energy coefficients, we find the feature vectors using Discrete Cosine Transform (DCT) analysis. Compute the MFCC feature vectors for the entire frame of the speech signal for the individual speaker. In order to avoid channel mismatch we used cepstral mean subtraction procedure for the entire utterance. Liftering is a procedure which is used to eliminate the effects of different roll off in various telephone channels on cepstral coefficients. In this work, the conventional speaker recognition system using MFCC feature will be the baseline system.

The new feature set used in our work is the Wavelet Octave Coefficients of Residues (WOCOR). A time-frequency vocal source feature extraction by pitch-synchronous wavelet transform, with which the pitch epochs, as well as their temporal variations within a pitch period and over consecutive periods can be effectively characterized [40]. The wavelet transform of time signal x(t) is given by equation (3).

$$w(a,\tau) = \frac{1}{\sqrt{|a|}}\int_t x(t)\psi^*\left(\frac{t-\tau}{a}\right) \qquad (3)$$

Where $\psi(t)$, a and τ are the mother wavelet function, scaling (or dilation) parameters and translation parameter respectively. Where $\psi\left(\frac{t-\tau}{a}\right)\frac{1}{\sqrt{a}}$ is named the baby wavelets. It is constructed from the mother wavelet by first, scaling $\psi(t)$ which means to compress or dilate $\psi(t)$ by parameter a and then moving the scaled wavelet to the time position of parameter τ . The compression or dilation of $\psi(t)$ will change the window length of wavelet function, thus changing the frequency resolution. Therefore, the ensemble of $\psi\left(\frac{t-\tau}{a}\right)\frac{1}{\sqrt{a}}$ constitutes the time-frequency building blocks of the wavelet transform [30]. The wavelet transform of discrete time signal x(n) is given by equation (4).

$$w(a,b) = \frac{1}{\sqrt{a}}\sum_n x(n)\psi^*\left(\frac{n-b}{a}\right) \qquad (4)$$

Where $a = \{2^k | k=1,2\ldots,K\}$ and $b = 1,2\ldots,N$, and N is the window length. $\psi^*(n)$ is the conjugate of the fourth-order Daubechies wavelet basis function $\psi(n)$. K=4 is selected such that the signal is decomposed into four sub-bands at different octave levels. At a specific sub-band, the time-varying characteristics within the analysis window are

measured as parameter b changes. To generate the feature parameters for pattern recognition, the wavelet coefficients with specific scaling parameters are grouped is given by equation (5).

$$W_k = \{ w(2k,b) | b = 1,2..N \} \quad (5)$$

where N is the window length. Each W_k is called an octave group. Then WOCOR parameters can be derived by using equation (6).

$$WOCOR_M = \left\{ \mathrm{P}W_k(m)\,\mathrm{P} \middle| \begin{matrix} m = 1,2..M \\ k = 1,2..4 \end{matrix} \right\} \quad (6)$$

where ||.|| denotes two-norm operation. Finally, for a given speech utterance, a sequence of $WOCOR_M$ feature vectors is obtained by pitch-synchronous analysis of the LP residual signal. Each feature vector consists of 4M components, which are expected to capture useful spectro-temporal characteristics of the residual signal. For each voiced speech portion, a sequence of LP residual signals of 30 ms long is obtained by inverse filtering the speech signal. The neighboring frames are concatenated to get the residual signal, and their amplitude normalized within (-1, 1) to reduce intra-speaker variation. Once the pitch periods estimated, pitch pulses in the residual signal are located. For each pitch pulse, pitch-synchronous wavelet analysis is applied with a Hamming window of two pitch periods long. For the windowed residual signal x(n) the wavelet transform is computed using equation (4).

2) Training Phase.

For speaker recognition, pattern generation is the process of generating speaker specific models with the collected data in the training stage. The mostly used modeling techniques for modeling include vector quantization [14] and Gaussian mixture modeling [15]. The VQ modeling involves clustering the feature vectors into several clusters and representing each cluster by its centroid vector for all the feature comparisons. The GMM modeling involves clustering the feature vectors into several clusters and representing all these clusters using a weighted mixture of several Gaussians. The parameters that include mean, variance and weight associated with each Gaussian are stored as models for all future comparisons. A GMM is similar to a VQ in that the mean of each Gaussian density can be regarded as a centroid among the codebook. However, unlike the VQ approach, which makes hard decision (only a single class is selected for feature vector) in pattern matching, the GMM makes a soft decision on mixture probability density function. This kind of soft decision is extremely useful for speech to cover the time variation.

In training phase, the first modeling technique we used in this work is Vector Quantization (VQ). After finding the MFCC feature vectors for the entire frame of the speech signal for the individual speaker, we have to find some of the code vectors for the entire training sequence with less number of code words and having the minimum mean square error. To find minimum mean square error with less number of code words by using VQ, we have two most popular methods namely K-means algorithm and Linde-Buzo and Gray (LBG) algorithms [23]. Vector quantization process is nothing but the idea of rounding towards the nearest integer.

The second modeling technique we used in our work, the Gaussian Mixture Modeling (GMM), which is most popular generative model in speaker recognition. The template models, VQ codebooks, can also be regarded as a generative model, although it does not model variations. The pattern matching can be formulated as measuring the probability density of an observation given the Gaussian. The likelihood of an input feature vectors given by a specific GMM is the weighted sum over the likelihoods of the M unimodal Gaussian densities [32], which is given by equation (7).

$$P(x_i | \lambda) = \sum_{j=1}^{M} w_j b(x_i | \lambda_j) \quad (7)$$

The likelihood of x_i given j^{th} Gaussian mixture is given by

$$b(x_i | \lambda_j) = \frac{1}{(2\pi)^{D/2} |\sum_j|} \exp\left\{ -\frac{1}{2} (xi - \mu j)^T \sum_j^{-1} (xi - \mu j) \right\} \quad (8)$$

Where D is the vector dimension, μ_j and Σ_j are the mean vectors and covariance matrices of the training vectors respectively. The mixture weights w_j are constructed to be positive and the sum to be one. The parameters of a GMM are: Mean (μ_j), Covariance (Σ_j) and Weights (w_j) can be estimated from the training feature vectors using the maximum likelihood criterion, via the iterative Expectation-Maximization (EM) algorithm (32). The next stage in the speaker recognition system will be the testing phase.

3) Testing Phase.

In this phase, feature vectors are generated from the input speech sample with same extraction techniques as in training phase. Pattern matching is the task of calculating the matching scores between the input feature vectors and the given models in recognition. The input features are compared with the claimed speaker pattern and a decision is made to accept or reject the claiming. Testing phase in the person authentication system includes matching and decision logic. The testing speech is also processed in a similar way and matched with the speaker models using Euclidean distance in case of VQ modeling and likelihood ratio in case of GMM modeling. Hence matching gives a score which represents how well the feature vectors are close to the claimed model. Decision will be taken on the basis of matching score, which depends on the threshold value.

The alternative is to employ verification through identification scheme. In this scheme the claimed identity model should give best match. The test speech compared with the claimed identity model, if it gives best match, then it is accepted as genuine speaker, otherwise, rejected as imposter.

For testing the performance of speaker recognition system, we have collected the speech database of students of SSIT at a sampling frequency of 8 kHz. Figure 1 shows speaker 1 sample speech signal of four sentences, which is collected by using microphone.

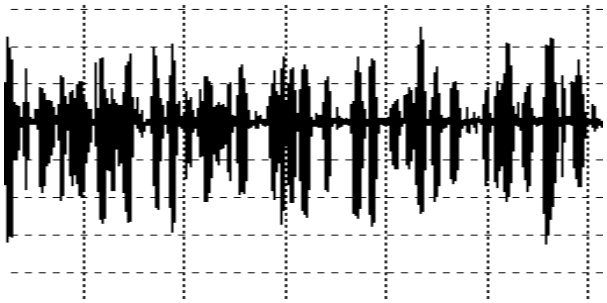

Figure 1. Sample of speech signals of speaker 1.

The SSIT database contains the speech data of 30 speakers, among them 20 were male and the remaining 10 were female. Four sentences are used for each speaker and 24 number of utterances of each sentence for each speaker. First 16 utterances are used for training and the remaining 8 utterances are used for testing. Table I shows the experimental results of different speaker identification and verification systems using SSIT speech database.

TABLE I. SPEAKER RECOGNITION SYTEM

Code book size	**MFCC-VQ based System**		
	Speaker Identification	***FAR***	***FRR***
32	98.75%	0%	0%
64	100%	0%	0%
Code book size	**WOCOR-VQ based System**		
	Speaker Identification	***FAR***	***FRR***
32	89.1667%	0.1293%	3.75%
64	96.25%	0.22%	1.624%
Gaussians	**MFCC-GMM based System**		
	Speaker Identification	***FAR***	***FRR***
32	100%	0%	0%
64	100%	0%	0%%
Gaussians	**WOCOR-GMM based System**		
	Speaker Identification	***FAR***	***FRR***
32	94.5833%	0.113%	3.333%
64	100%	0%	0%%

The performance of the conventional MFCC-VQ based speaker recognition system with code book size of 64 gives better result compared to WOCOR-VQ based system. The WOCOR-GMM and MFCC-GMM based speaker recognition systems with 64 Gaussians also gives better result. Finally, we combine the matching scores of MFCC-VQ and WOCOR-GMM based systems. These combined system become one of the baseline system, which is used for developing the multimodal system for person authentication. The reasons for the same are the different feature extraction and modeling techniques are used.

B. Signature Recognition System

Signature recognition is the task of recognizing signatories by using their signatures. Signature is a behavioral biometric, the features of signature are variant with respect to time and the forgers can easily fool the system by reproducing the signatures of the correct persons. Irrespective of the above limitations we can still use signature as our best biometric feature, since the signature is a unique identity of an individual and is being used extensively in practical systems. No two signatures can be identical, unless one of them is a forgery or copy of the other [35]. The signature recognition systems find applications in government, legal and commercial areas. Signature verification is the verification of given signature of claimed identity of a person. There are two types of signature verification systems in practice, namely, online and offline [17], [18]. Online signature verification uses information collected dynamically at the time of signature acquisition like timing, acceleration, velocity, pressure intensity and also termed as dynamic signature verification. Offline signature verification uses only the scanned image of signature and also termed as static signature verification.

In case of online signature verification during the training phase, the user supplies a set of reference signatures measured in terms of dynamic features mentioned above. These dynamic features along with signatures are stored as reference templates. When a test signature is input to the system in terms of these dynamic features, it is compared to each of the reference signatures of the claimed person. Based on the resulting comparison distance, the claimed identity is either accepted or rejected. Most of the existing signature verification systems are based on online approach. Not much importance has been given to the offline signature verification as it is relatively complex. The complexity may be due to the two dimensional nature of offline signature compared to one dimensional online signature. However, once we have signature images, then we can view signature verification as a pattern recognition problem. The online and offline approaches exploit different aspect of signature information for verification, namely, dynamic and static. The development of offline signature verification and integrating with existing online system may provide improved performance as well as robustness. It is therefore aimed to develop offline signature verification system static features like aspect ratio, horizontal projection profile, vertical projection profile and discrete cosine transform features.

1) Feature extraction phase.

Feature extraction plays a very important role in offline signature verification. Unlike our speaker recognition case, we are not going model the feature vectors up to some codebook level. Here feature vectors itself will give the training sequence. In this work the features of signature are extracted by using Discrete Cosine Transform (DCT) analysis, Vertical Projection Profile (VPP) analysis and Horizontal Projection Profile (HPP) analysis. The VPP and HPP are static features of a signature and DCT is a global feature of a signature image. Since our signature is an image, it will have the gray levels from 0 to 255 and to compute the maximum gray level the histograms of all images are used. VPP and HPP are the kind of histograms. VPP gives the horizontal starting and ending points and HPP gives the vertical starting and ending points of the image. The size of VPP and HPP is equal to the number of columns and the number of rows in the signature image respectively. Since, the size of signature regions are not

constant even for a single user, in this work we are taking average value of vertical projection profile as a feature.

$$vpp_{avg} = \frac{1}{N}\sum_{q=1}^{N}\sum_{p=1}^{M} A(p,q) \quad (9)$$

$$hpp_{avg} = \frac{1}{M}\sum_{p=1}^{M}\sum_{q=1}^{N} A(p,q) \quad (10)$$

The signature image intensity A (p, q) at p^{th} row and q^{th} column indices respectively. Where M is number of rows in an image and N is number of columns in image. Equation (11) gives the DCT coefficient corresponding to p^{th} row and q^{th} column of an input signature image.

$$B_{pq} = \alpha_p \alpha_q \sum_{m=0}^{M-1}\sum_{n=0}^{N-1} A_{mn} \cos\left(\frac{\pi(2m+1)p}{2M}\right)\cos\left(\frac{\pi(2n+1)q}{2N}\right) \quad (11)$$

where $\alpha_p = \left\langle \frac{1}{\sqrt{M}} \right.$ for p=0 and $\alpha_p = \left\langle \sqrt{\frac{2}{M}} \right.$ for $1 \le p \le (M-1)$

$\alpha_q = \left\langle \frac{1}{\sqrt{N}} \right.$ for q=0 and $\alpha_q = \left\langle \sqrt{\frac{2}{N}} \right.$ for $1 \le q \le (N-1)$

The performance of signature recognition system depends on the way in which the DCT coefficients are considered. The zonal coding of DCT coefficients of signature image are used for better performance, which gives concentration at low spatial frequencies.

2) Testing Phase.

For the identification or verification, same set of features which have been extracted during registration process are extracted from the input samples scanned or recorded using input devices like writing pads, to form the feature vectors. Verification is 1 to 1 matching while identification is 1 to n matching. In verification, the individual claims his/her identity which is verified by comparing these features vectors by the feature vectors of the individual which he/she claimed to be. If the matching score crosses the predefined threshold then the system verifies the individual as authentic user. In identification, the feature vectors of the individual are compared with the feature vectors of every individual stored in the database. If the highest matching score crosses the predefined threshold, then it identifies the individual as the person whose matching score is the highest otherwise the system suggest few top most matches. The matching algorithm is needed to compare the samples and computes the matching score and decide if two samples belong to the same individual or not by comparing the matching score against the acceptance threshold. However, it is possible that sometimes the output of a biometrics system may be wrong. Therefore, the performance of a biometrics system is measured in terms of two errors: FAR and FRR. In order to design the multimodal system using speech and signature features, we have collected the signature database from the same 30 students who had given their speech samples while collecting the speech database. For every writer we have taken 24 samples of signatures and scanned them by using HP Scan jet 5300C scanner at 300 digits per inch resolution and stored them in 'bmp' format. After scanning the signatures, we have cropped all the 24 signatures of individual writer by using Windows Picture manager. Figure 2 shows one of the sample signature of user 1.

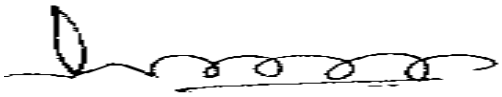

Figure 2. Sample signature of user 1.

During the training session, we considered the first 16 signatures of each writer and extract the features from those signatures by using VPP, HPP and DCT analysis. The three feature models are obtained for all 30 users. In testing phase, we have used the remaining 8 signatures for each writer. For the given test signature, we have to extracted the VPP, HPP values and DCT coefficients separately by VPP-HPP-DCT analysis. After getting these values, we found the minimum distance between the VPP-HPP-DCT values and the feature vectors of all the writers corresponding to each of the model. Table II shows the performance of different signature recognition systems using SSIT signature database. The VPP-HPP-DCT method gives highest performance (86.66%) compared to the other systems.

TABLE II. SIGNATURE RECOGNITION SYTEM

System	VPP-HPP-DCT based System
	Signature Identification
VPP	21.25%
HPP	30.4167%
VPP-HPP	56.25%
DCT	72.9167%
VPP-HPP-DCT	86.667%

To improve the performance of the signature recognition system, along with the baseline VPP-HPP system the DCT coefficients are used. A modified system uses VPP and HPP vectors with Dynamic Time Warping (DTW) for the optimal cost. DTW is a pattern matching technique which aims at finding the minimum cost path between the two sequences having different lengths [26]. A very general approach to find distance between two time series of different sizes is to resample one of the sequence and comparing the sample by sample. The drawback of this method is that there is a chance of comparing the samples that might not correspond well. This means that comparison of two signals correspond well when there is a matching between troughs and crests. DTW solves this method by considering the samples with optimum alignment. The DTW computation starts with the warping of the indices of two sequences. The two sequences are compared with some distance measures like Euclidean distance at each and every point, so as to obtain the distance matrix. These distances in the matrix are termed as local distances. Let the Matrix be D and the sequences are A, B with lengths M, N respectively. Then D is calculated using equation (12).

$$D(i, j) = \text{distance}(A(i), B(j)) \quad (12)$$

where i varies from 1 to M and j varies from 1 to N. The distance here considered is Euclidean distance. The modified

feature vectors obtained from the signature image A(i,j) of size MxN are given in the equations (13) and (14).

$$vpp_{(j)} = \sum_{i=1}^{M} A(i, j) \text{ where j=1,2,3……. N} \quad (13)$$

$$hpp_{(j)} = \sum_{j=1}^{N} A(i, j) \text{ where i=1,2,3….. M} \quad (14)$$

Calculate the DTW distance values separately for VPP and HPP vectors from all the users for all the training images to the testing image and obtain distances from each user using average distance method. Normalize each of the distance of a particular feature using one of the normalization methods and use sum rule for fusion of match scores obtained using each model. Assign the test signature to the user who produces least distance in fused sum vector. Table III shows the results of signature verification system using SSIT signature database.

TABLE III. SIGNATURE VERIFICATION SYTEM

System	VPP-HPP-DCT based System		
	FAR	***FRR***	***Average Error***
VPP-HPP-DCT	1.2931%	37.5%	19.396%
Modified VPP-HPP-DCT	0.1149%	3.333%	1.7241%

The VPP-HPP-DCT based signature identification gives better result compare to other systems. The modified VPP-HPP-DCT system gives even better in signature verification. These systems are used in multimodal biometric system for identification and verification of test signature respectively.

C. Handwriting Recognition System

Handwriting biometric feature can also be used for person authentication [25]. Most of the existing works on handwriting information is for forensic investigation. The scope includes identifying the author of the given handwritten script from the group of available large population. The end result may be a subgroup of most likely population. This subgroup may then be carefully analyzed by the human experts to identity the correct person who might have written the script. Thus using handwriting information in criminal investigation is an age old method. Handwriting biometric feature may also posses several characteristics to qualify it for use in person authentication. Relatively few works have been done in this direction [25]. With the integration of pen-based input devices in PDA and Tablet PCs strongly advocates the use of handwriting information for person authentication due to ease of collection. Handwriting verification can also be done either in online or offline mode as in signature verification. Online handwriting verification exploits similar dynamic features as in signature verification. Thus it is easy to extend the online signature verification approach to handwriting verification. Initially an online handwriting verification system will be developed. However, it should be noted that there is a significant difference between signature and handwriting. Signature is one pattern from hand, but it will not use any language specific information. Alternatively, handwriting exploits language information at various levels, starting from character set. Since this has been trained during initial days of learning stage of language, it is possible that, we may find more regular and reliable feature from handwriting for person authentication. The offline handwriting verification can be approached by using the information from the offline signature verification literature and also from the speaker recognition literature. An offline system is initially developed using the technique developed for offline signature verification. Later techniques available in speaker recognition literature can be mapped here to further improve the performance or develop a new technique for verification. For instance, well known text dependent speaker verification technique using Dynamic Time Warping (DTW) can be extended to offline handwriting verification in the text dependent mode using VPP features. Finally a hybrid handwriting verification system using offline and online approaches may be developed to provide improved performance and robustness.

The methods in handwriting biometrics account for both offline and online with verification and identification modes. The two recent approaches for handwriting recognition which have proved fruitful are based on a textural feature, whereas the second method zooms in on character shape elements [29]. The first method refers to angles and curvature in handwriting which are determined by the degrees of freedom in wrist and finger movement, which in turn depend on the pen grip attitude and the applied grip forces. The other approaches include use of Hidden Markov Models, Gray level distribution [28], Support vector machine, and connected component contours [27]. The Gray level distribution approach is used for handwriting recognition.

1) Feature extraction phase.

Mainly dynamic time warping in context of images is used for word matching which uses vectors like normalized upper word and lower profile, back ground ink transitions etc.,. The features from the handwriting image considered in our work are VPP vector and HPP vector. The VPP is an array that contains sum of gray levels of each column in a handwriting image. This feature signifies the variations of Gray level distribution along the length of the image. This VPP vector is a unique feature for a given user and will vary from user to user. Even the same user will have variations. The important and the uniqueness of the information present in the HPP vectors are equally important as that of VPP vectors. So along with the VPP vector extraction, another feature HPP vector is obtained from the handwriting image. This HPP vector gives the information about the variations of the handwriting along the lateral extent. The handwriting recognition system runs on the same lines as of the signature recognition system.

2) Testing Phase.

Writer recognition system is built using the individual words, segmented from the sentence considered for handwriting and combined later for better performance. In order to obtain a correct segmentation, a threshold is calculated that distinguish words and characters. After obtaining the threshold, words are segmented by obtaining the VPP vector and examining its intensity profile. The each word extracted from the sentence now act as the images to be tested. The algorithm proposed for a full sentence is applied for each word. First obtain the image from which words should be segmented out and let N numbers of words are segmented.

Consider one word and apply already proposed algorithm for all sentence. Obtain the DTW distances from each user by averaging method. Next, normalize the distances and repeat the same procedure for all the words. The normalized distances are fused using the fusion principle. Obtain minimum distance and its corresponding user there by identifying the user. At the fusion level, distances are fused using sum rule. The similar procedure is used for finding the HPP vectors.

For handwriting recognition system we created database of same 30 users of speech and signature recognition systems. The same sentence used for verification for all users. The sentence is written on A4 sheet with eight equal rectangular boxes. This sheet was scanned using HP scanner, at a resolution of 300 digits per inch. Then sentences are separated out and stored in bits mapping format. In our experiments, five samples are used for training and three samples are used for testing. Figure 3 shows the sample of handwriting of user 1.

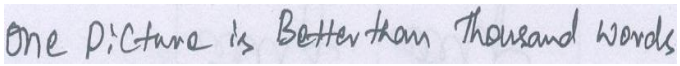

Figure 3. Sample of handwriting of user 1.

Table IV shows the performance of handwriting identification and verification system using SSIT database.

TABLE IV. HANDWRITTING RECOGNITION SYTEM

System	VPP-HPP-based System		
	Identification	***FAR***	***FRR***
HPP	80.63%	1.461%	26.6%
VPP	90.02%	0.062%	0.362%

The VPP based handwriting recognition system gives better result compared to HPP based system. The reason is that, the richest image information obtained from Gray level distribution along the length of the handwriting image. The HPP vectors are equally important as that of VPP vectors which gives the information about the variations of the handwriting along the lateral extent. The combined feature gives the complete behavior of handwriting image of a user, hence VPP-HPP based system is one the unimodal system in our multimodal biometric person authentication.

III. MULTIMODAL BIOMETRIC PERSON AUTHENTICATION SYSTEM

A. Development of Multimodal System

Multimodal or Multi-biometric systems, remove some of the drawbacks of the unimodal systems by grouping the multiple sources of information. These systems utilize more than one physiological or behavioral characteristic for enrollment and identification. Once the unimodal systems are developed, then the next step is to develop multimodal system by integrating them suitably. The unimodal systems using speech, signature and handwriting information are ranked according to their performance. Based on this, the best performing system is used as the baseline system to which other systems are integrated. The integration can be done at any of the following three levels: feature, measurement and score levels [2]. A tight integration is possible if it is done at the feature level. However, the difficulty associated is the different nature of features and also significant variation of person information. This difficulty can be overcome by integrating at the measurement level, but the level of person information present may be smoothened out to some extent due to modeling. To that extent the level of coupling will be loose or moderate. The integration of measurement values may also harm the combined system, if one of the systems provides poor performance. Under such condition, the safe way to integrate is at the score level. This level of fusion is immune to any poor performance, since already decision is made about the person. But the amount of improvement achieved after combination may be relatively low. For the selected baseline system, only the next best performing unimodal system is integrated at the feature level. One approach for the same is to extract same features say, Discrete Cosine Transform (DCT) values and normalize them suitably and combine them. Alternatively, the features can be applied to one more level of smoothing using feature modeling techniques to obtain modified features that are similar for both biometric features. These features are then used for modeling. As a result of this a bimodal person authentication system is developed by integrating at the feature level. At the next level, the best performing unimodal system is integrated with the feature level integrated bimodal system. The integration is done at the measurement level. The measurement scores from the two systems are suitably normalized combined and evaluated [36]. This results in tri-modal biometric person authentication system. At the final level, the unimodal system based on the fourth biometric feature is integrated into the tri-modal system at the score level. For this several combination techniques are explored to obtain maximum gain. This will result in the development of the multimodal biometric person authentication system using all the possible biometric features.

The particular biometric feature selection for developing unimodal system needs to satisfy different characteristics, as we mentioned above. However, some of the parameters are observed and studied during the design of unimodal systems. The following are the certain parameters to decide whether a biometric trait can be used for person authentication or not.

They are: (1) how common the trait is found in individual, (2) how much the trait varies from individual to individual, (3) how the trait varies with the age of the individual, (4) how easily the trait be collected, (5) how easily the trait can be processed and how is the accuracy and speed of the system built using the trait, (6) how people adopt the technology in their day to day life. The speech, signature and handwriting biometrics are fulfill the above requirements and characteristics of biometric person authentication. By combining the offline signature recognition, offline handwriting recognition and the text dependent speaker recognition systems, the challenge-response type of authentication can be facilitated.

With these factors, the three best performing unimodal systems are combined using score level fusion. Since we are using score level fusion, there are no special steps involved in the training of biometric system. In the score level fusion, scores obtained at the output of the classifier are fused using

some rules. The simple rules of fusion are Sum rule, Product rule, Min Rule, Max rule and Median rule. The Sum rule and Product rule assume the statistical independence of scores from the different representations [6]. The outputs on the individual matchers need not be on the same numerical scale. Due to these reasons, score normalization is essential to transform the scores of the individual matchers into a common domain prior to combining them. Score normalization is a critical part in the design of a combination scheme for matching score level fusion. Min-max and Z-Score normalization are the most popular techniques used for normalization. The present work uses the Z-Score normalization techniques for the individual matcher and the Sum rule for integrating the normalized scores. Using these two principle techniques, the multimodal biometric system is designed using three unimodal systems. Once the multimodal system is developed the next stage is performance and robustness evaluation.

B. Performance and Robustness Evaluation

There are standard databases for the individual evaluation of the unimodal biometric systems, like YOHO database, IITG database etc... However, such an evaluation is only for finding the performance of the particular unimodal system in an absolute sense. To have comparative study evaluate the strength of multimodal system on common platform, it is proposed to develop a multimodal database for these three biometric features. For this reason we have prepared our own SSIT database of 30 users.

The database consists of 24 samples of speech information, 24 samples of signature and 8 samples of handwriting for each user. Once the database is developed, then the performance is evaluated first for each of the unimodal systems. The performance is then evaluated for multimodal system using all the three features. Such evaluation provides a systematic comparison between unimodal and multimodal systems. The main features considered in developing a multimodal system are handwriting, signature and speech. The following are the steps involved in the implementation of multimodal biometric person authentication system based on unimodal system performance.

a) Collect the individual matching scores of the unimodal systems for every user.

b) Normalize the matching scores using normalization techniques and integrate the scores by using fusion rules.

c) Assign the multiple biometric to a particular person who produces the minimum score.

Table V shows results of different multimodal systems based on the combination of different unimodal systems.

TABLE V. PERFORMANCE OF MULTIMODAL BIOMETRIC SYSTEMS

System	VPP-HPP-based System	
	FAR	***FRR***
VQ-MFCC for Speech VPP-HPP-DCT for Signature VPP-HPP with DTW for Hadwriting	0%	0%
VQ-WOCOR for Speech VPP-HPP-DCT for Signature VPP-HPP with DTW for Hadwriting	0%	0%
GMM-MFCC for Speech VPP-HPP-DCT for Signature VPP-HPP with DTW for Hadwriting	0%	0%
GMM-WOCOR for Speech VPP-HPP-DCT for Signature VPP-HPP with DTW for Hadwriting	0%	0%

Table V proves the advantages of multimodal biometric system through its performance and robustness evaluation by using more number of biometrics for person authentication. The other major factors to be concentrated along with the development of multimodal system are the fusion rules and the normalization techniques. The score level fusion technique with Sum rule is employed in all the cases. The normalization techniques are used for the maintenance of the homogeneity among the scores obtained from different features. As a result, we are using the three possible combinations of unimodal systems to develop four multimodal systems. The each multimodal system identification performance is 100% and the verification performance is 0% error rates, even though there are some error rates in respective unimodal system.

IV. CONCLUSION AND FUTURE WORK

The trend of multimodal biometrics is spreading for the authentication process to maintain the interests regarding the security as strong as possible. The vital features that encourage the use of multimodal biometrics are the performance and accuracy along with the ability to overweigh the drawbacks of unimodal biometric systems. In this work we demonstrated multimodal biometric person authentication system using three biometric features. We generated our own database of 30 users and effectively using the principle of matching score fusion and normalization technique for developing multimodal system. Further, we combined the multimodal systems shown in Table V using normalization and fusion techniques. This system gives the identification performance is 100% and the verification performance is 0%, in terms of FAR 0% and FRR is 0%. As a result, we implemented multimodal biometric person authentication system using speech, signature and handwriting features which provides 0% error rates.

The future work may include integrate the biometric features at the feature level for improving the performance.

Further, to make the biometric person authentication system more practical: use more number of users, different sessions of collecting data of the same users, and multiple sensors for data collection. Also, by combining the offline system with online system may improve the performance.

ACKNOWLEDGMENT

I would like to thank my students for providing their speech, signature and handwriting information in generating the multimodal SSIT database.

REFERENCES

[1] L.Gorman, "Comparing passwords, tokens and biometrics for user authentication," IEEE Proc., vol. 91, no.12, Dec. 2003.

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

[3] R. Bruneelli and D.Falavigna, "Person identificationusing multiple cues," IEEE Trans. PAMI, vol. 17, no.10, pp.955-966, oct.1995.

[4] L. Hong and A.K. Jain, "Integrating faces and fingerprints for person identification," IEEE PAMI, vol. 20, no. 12, pp. 1295-1307, Dec. 1998.

[5] V. Ghattis, A.G. Bors and I. Pitas, " Multimodal decision level fusion for person authentication," IEEE Trans. Systems, Man and Cybernatics, vol. 29, no. 6, pp, 674-680, Nov. 1999.

[6] R. W.. Frischholz and U. dieckmann, "Bioid: A multimodal biometric identification system," IEEE Computer Society, pp. 64-68, Feb.2000. Name Stand. Abbrev., in press.

[7] A. Kumar et. al., "Person verification using palmprint and handgeometry biometric," proc. Fourth Int. Conf. AVBPA, pp.668-678, 2003.

[8] S.Ribaric, D. Ribaric and N. Pavesic, "Multimodal biometric user identification system for network based applications," IEEE Proc. Vision, Image and Signal Processing, vol. 150, no.6, pp.409-416, 2003.

[9] A.K. Jain and Ross, "Learning user specific parameters in multibiometric system," Proc. Int. Conf. Image Processing (ICIP), pp. 57-60,2002.

[10] A. K.Jain, L.Hong and Y. Kulkarni, " A multimodal biometric system using fingerprint, face and speech," Proc. Second Int. Conf. AVBPA, pp.182-187, 1999.

[11] S.Ribaric, I. Fratic and K. Kris, " A biometric verification system based on the fusion of palmprint and face features," Proc. Fourth Int. Symposium Image and Signal Processing, pp. 12-17, 2005.

[12] B.Duc et. al., "Fusion of audio and video information for multimodal person authentication," Pattern Recogntion Letters, vol. 18, pp.835-845, 1997.

[13] S. Furu, "Cepstral analysis techniques for automatic speaker verification," IEEE Trans, Acoust., Speech Signal Processing, vol. 29(2), pp.254-272, 1981.

[14] F.K.Soong, A.E.Rosenberg, L.R. Rabiner and B.H. Jvang," A Vector quantization approach to speaker recognition," Proc., IEEE, Int., Conf., Acoust.,

Speech Signal Processing, vol. 10, pp.387-390, Apr.1985.

[15] D.A.Reynolds, "Speaker identification and verification using gaussian mixture speaker models," Speech Communication, vol. 17, no.1-2, pp.91-108, 1995.

[16] S.R.M. Prasana, C.S. Gupta, and B. Yegnanarayana, "Extraction of speaker-specific excitation information from linear prediction residual of speech," Speech Communication, vol. 48, no.10, pp.1243-1261, oct.2006.

[17] B.S.Atal, "automatic recogntion of speakers from their voices," IEEE Proc. vol. 64(4), pp.460-475, Apr. 1976.

[18] A.Eriksson and P.Wretling, "How flexible is the Human Voice? Acase study of Mimicry," Proc. European Conf. on Speech Tech. Rhodes, 1043-1046, 1997.

[19] V.S.Nalwa, "Automatic online signature verification," Proc. IEEE, vol. 85, no.2, pp.213-239, Feb. 1997.

[20] W.Hou, X.Ye and K.Wang, "A survey of offline signature verification," Proc.Int. Conf. Intelligent Mechatronics and Automation, pp. 536-541, Aug.2004.

[21] T. Scheidat, c. Vielhauer and J. dittmann, "Single-semantic multi-instance fusion of handwritting based biometric authentication systems," Proc.Int. Conf. IEEE-ICIP, pp. II-393-II-396, 2007.

[22] A.Rosenberg, "Automatic speaker verification: a review," Speech Communication, vol. 17, no.1-2, pp.91-108, 1995.

[23] Y.Linde, A.Buzo, and R.M.Gray, " An algorithm fot vector quantizer design," IEEE Trans. on Communications, vol. COM_28(1), pp.84-96, Jan. 1980.

[24] Chaur-Heh Hsieh, "DCT based code book design for vector quantization of images," IEEE Trans. Systems for Video Technology, vol. 2, no.4, pp.401-409, Dec 1992.

[25] F.Ramann C.Vielhuue, and R. Steinmetz, "Biometrics applications based on handwriting," IEEE Proc. Int. Conf. on Multimedia and Expo, ICME, vol2, pp.573-576, 2002.

[26] T.M. Mat and R. Manmatha, "Word image matching using dynamic time warping," Proc. IEEE, Computer Vision and Pattern Recgn., vol.2, pp.521-527, June 2003.

[27] L.Schomaker and M. Bulacu, "Automatic writer identification using connected component contours and edge-based features of uppercase western sript," IEEE Tran. Pattern Analysis and Machine Intelligence, vol. 26, pp.787-789, 2004.

[28] M.Wirotius, A. Seropian, and N. Vincent, "Writer identification from Gray level distribution," Proc. 7th Int. Conf. on Document Analysis and Recognition, ICDAR, pp.1168-11721, 2003.

[29] F. Nobard, "Handwritten signature verification: Global approach, Fundamentals in handwriting recognition," Springer-verlag, berlin series: Computer and System Science, 124, pp.445-495, 1991.

[30] Y.T. Chan. Wavelet Basics. Kluwer Academic Publishers Group, 1996.

[31] G. strang and T. Nguyen. Wavelets and Filter Banks. WellesleyCambridge Press, 1996.

[32] A. Sanker and C.H. Lee, "A maximum-Likelihood approach to stochastic matching for robust speech recognition," IEEE Trans. Speech-Audio Processing, 4(3): 190-202, 1996.

[33] L.E. Baum and T. Petie, "Stastical inference for probabilistic functions of finite state Markov chains," Ann. Mat. Stat., 37, pp.1554-1563, 1966.

[34] F.Leclerc and R. Plamodon, "Automatic Signature Verification," Int. Jr. of Patt. Recogn. and Art. Intelligence, vol. 18, no. 3, pp. 643-660, 1994.

[35] M.Ammar, Y.Yoshido, and T.Fukumura, "Structural description and classification of signature images," Patt. Recogn. vol. 23, no. 7, pp. 697-710, 1990.

[36] A. Jain, K.Nandakumar, and A. Ross, "Score normalization in multimodal biometric systems," Elsevier, Patt. Recogn. Journal, vol. 38, pp. 2270-2285, Jan. 2005.

[37] I.Daubechies, "Ten Lectures on Wavelets," Phaladelphia, PA: Siam, pp.36-106, 1992.

[38] Earl Gose, Richard Johnsonbaugh, and Steve Jost, Pattern Recognition and Image Analysis, PHI: Pentice Hall publisher, pp. 329-409, 1997.

[39] Meenakhsi, Sargur Srihari, and aihuxu, "Offline signature verification and identification using distance measures," Int. Journal, Patt. Recogn. and Art. Intelligence, vol. 18, no.7, pp. 1339-1360, 2004.

[40] Nengheng Zheng, Tan Lee and P.C. Ching, "Integration of complementary Acoustic Features for Speaker Recognition," IEEE Signal Processing Letters, vol. 14, no.3, March 1997.