Combined features for an Secured Authentication system

Mahesh P.K. #1 and M.N. Shanmukha Swamy *2

Department of Electronics and Communication,
J.S.S. research foundation, Mysore University, S.J.C.E., Mysore, Karnataka, India
mahesh24pk@gmail.com

* *Department of Electronics and Communication,*
J.S.S. research foundation, Mysore University, S.J.C.E.,Mysore, Karnataka, India
mnsjce@gmail.com

***Abstract*—Biometrics based personal identification is regarded as an effective method for automatically recognizing, with a high confidence, a person's identity. This paper proposes the multimodal biometrics system for identity verification using two traits, i.e., speech signal and palmprint. The proposed system is designed for applications where the training data contains a speech signal and palmprint. It is well known that the performance of person authentication using only speech signal or palmprint is deteriorated by feature changes with time. Integrating the palmprint and speech information increases robustness of person authentication. The final decision is made by fusion at matching score level architecture in which feature vectors are created independently for query measures and are then compared to the enrolment templates, which are stored during database preparation. Multimodal system is developed through fusion of speech signal and palmprint recognition.**

Keywords –Biometrics; multimodal; speech signal; palmprint; fusion; matching score.

I. INTRODUCTION

Unimodel biometric systems, relying on the evidence of a single source of biometric information for authentication, have been successfully used in many different application contexts, such as airports, passports, access control, etc. However, a single biometric feature sometimes fails to be exact enough for verifying the identity of a person. By combining multiple modalities enhanced performance reliability could be achieved. Due to its promising applications as well as the theoretical challenges, multimodal biometrics has drawn more and more attention in recent years [1]. Although information fusion in a multimodal system can be performed at various levels, integration at the matching score level is the most common approach due to the case in accessing and combining the score generates by different matchers. Since the matching scores output by the various modalities are heterogeneous, score normalization is needed to transform these scores into a common domain, prior to combining them. Speech signal and palmprint multimodal biometrics are advantageous due to the use of non-invasive and low-cost acquisition.
Multimodal systems also provide anti-spooling measures by making it difficult for an intruder to spool multiple biometric traits simultaneously. However, an integration scheme is required to fuse the information presented by the individual modalities.

The paper presents a novel fusion strategy for personal identification using speech signal and palmprint biometrics [2] at the feature level fusion scheme. The proposed paper shows that integration of speech signal and palmprint biometrics can achieve higher performance that may not be possible using a single biometric indicator alone. 2D Gabor filter with Hamming distance and Mel Frequency Cepstral Coefficients (MFCC) with Gaussian Mixture Model (GMM) are used for feature vector fusion context for palmprint and speech signal respectively.

The rest of this paper is organized as fallows. Section 2 presents the system structure, which is used to increase recognition quality. Section 3 presents feature extraction using 2D Gabor and MFCC. Section 4, the individual traits are fused at matching score level using weighted sum of score technique. Finally, the experimental results are given in section 5. Conclusions are given in the last section.

II. SYSTEM STRUCTURE

The multimodal biometric system is developed using two traits i.e. speech signal and palmprint as shown in Fig. 1. For

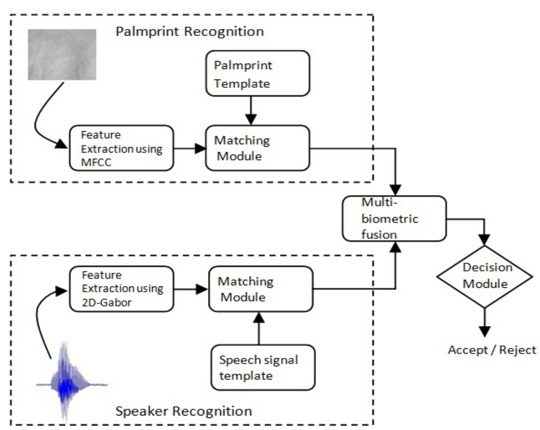

Fig. 1 Block diagram of speech signal and palmprint multimodal biometric system

the speech signal and palmprint Recognition, the input image is recognized using MFCC and 2D Gabor filter algorithm respectively. When we are using a gabor filter, the matching score is calculated using Hamming distance also when we are using MFCC, GMM is used. The modules based on the individual traits returns an integer vector after matching the database and query feature vectors. The final score is generated by using sum of score technique using FAR and FRR at matching score level, which is passed to the decision module.

III. Feature Extraction Using MFCC and Gabor Filter

A. *Feature Extraction Using MFCC*

Feature extraction is the first component in an automatic speaker recognition system [3]. This phase consists of transforming the speech signal in a set of feature vectors called also parameters. The aim of this transformation is to obtain a new representation, which is more compact, less redundant, and more suitable for statistical modeling and calculation of distances. Most of the speech parameterizations used in speaker recognition systems relies on a Cepstral representation of the speech signal [4].

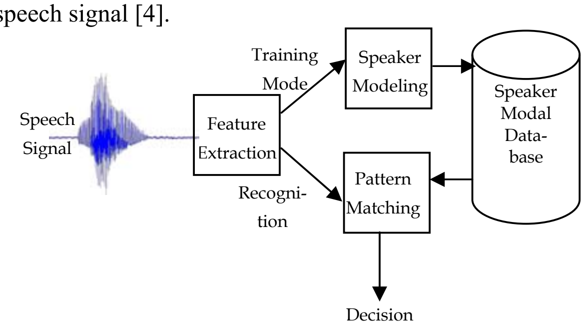

Fig.2 Componets of a speaker recognition system

The Mel-frequency Cepstral coefficients (MFCC) are motivated by studies of the human peripheral auditory system. Firstly, the speech signal x(n) is divided into Q short time windows which are converted into the spectral domain by a Discrete Fourier Trans form(DFT). The magnitude spectrum of each time window is then smoothed by a bank of triangular bandpass filters (Figure 3) that emulate the critical band processing of the human ear.

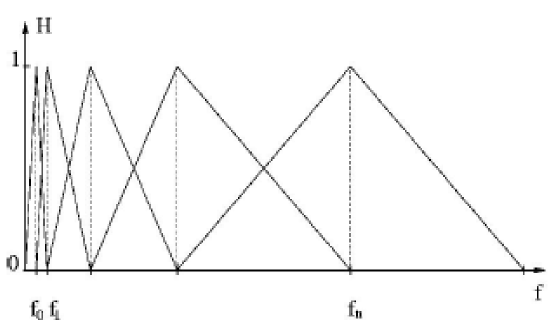

Fig.3 Mel filter bank

Each one of the bandpass filter H (k, m) computes a weighted average of that subband, which is then log|.| arithmically compressed:

$$X^1(m) = \ln\left(\sum_{k=0}^{N-1}\left|X(k)\right|H(k,m)\right) \tag{1}$$

where *X (k)* is the DFT of a time window of the signal *x(n)* having the length *N,* the index *k, k =0, . . . , N − 1,* corresponds to the frequency $f_k = k\, f_s / N$, with f_s the sampling frequency, the index m, m =1, . . . M and M << N, is the filter number, and the filters H (k, m) are triangular filters defined by the center frequencies *fc (m)* (Sigurdsson et al., 2006). The log compressed filter outputs X f (m) are then decorrelated by using the Discrete Cosine Transform (DCT):

$$c(l) = \sum_{m=1}^{M} X'(m)\cos\left(l\frac{\pi}{M}\left(m-\frac{1}{2}\right)\right) \tag{2}$$

where $c(l)$ is the l^{th} MFCC of the considered time window.

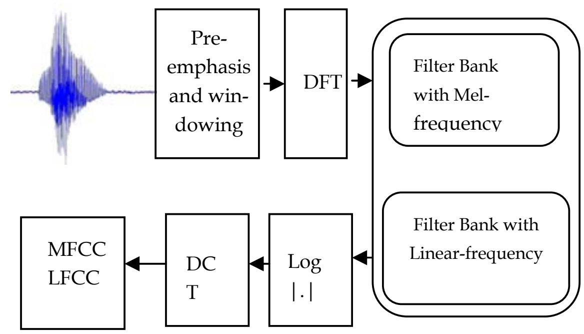

Fig.4 Extraction of MFCC and LFCC parameter

There are several analytic formulae for the Mel scale used to compute the center frequencies *fc (m)*. In this study we use the following common mapping:

$$B(f) = 2595\log_{10}\left(1+\frac{f}{700}\right) \tag{3}$$

B. *The Gaussian Mixture Model*

In this study, a Gaussian Mixture Model approach proposed in [5] is used where speakers are modeled as a mixture of Gaussian densities. The use of this model is motivated by the interpretation that the Gaussian components represent some general speaker-dependent spectral shapes and the capability of Gaussian mixtures to model arbitrary densities.

The Gausssian Mixture Model is a linear combination of M Gaussian mixture densities, and given by the equation,

$$p(\vec{x}\mid\lambda) = \sum_{i=1}^{M} p_i b_i(\vec{x}) \tag{4}$$

Where $\vec{x}$ is a D-dimensional random vector,

$b_i(\vec{x}), i = 1,...M$ are the component densities and p_i, i=1,…M are the mixture weights. Each component density is a D-dimensional Gaussian function of the form

$$bi(\vec{x}) = \frac{1}{(2\Pi)^{D/2} |\sum_i|^{1/2}} \exp\left\{-\frac{1}{2}(\vec{x}-\vec{\mu_i})^T \sum_i^{-1}(\vec{x}-\vec{\mu_i})\right\} \quad (5)$$

Where $\vec{\mu_i}$ denotes the mean vector and Σ_i denotes the covariance matrix. The mixture weights satisfy the law of total probability, $\sum_{i=1}^{M} p_i = 1$. The major advantage of this representation of speaker models is the mathematical tractibility where the complete Gaussian mixture density is represented by only the mean vectors, covariance matrices and mixture weights from all component densities.

C. Feature Extraction and Coding (Gabor Filter)

We proposed a 2D Gabor phase coding scheme for palmprint representation[6]. The circular Gabor filter is an effective tool for texture analysis, and has the following general form.

$$G(x,y,\theta,u,\sigma,) = \frac{1}{2\pi\sigma^2}\exp-\left\{\frac{x^2-y^2}{2\sigma^2}\right\}\exp\{2ni(ux\cos\theta + uy\sin\theta)\} \quad (6)$$

Where $i = \sqrt{-1}$, u is the frequency of the sinusoidal wave, θ controls the orientation of the function, and σ is the standard deviation of the Gaussian envelope. To make it more robust against brightness, a discrete Gabor filter, $G(x,y,\theta,u,\sigma,)$, is turned to zero DC(direct current) with the application of the following formula:

$$\tilde{G}[s,y,\theta,u,\sigma] = G[x,y,\theta,u,\sigma]\frac{-\sum_{i=-n}^{n}\sum_{j=-n}^{n} G[i,y,\theta,u,\sigma]}{(2n+1)^2} \quad (7)$$

Where $(2n + 1)^2$ is the size of the filter. In fact, the imaginary part of the Gabor filter automatically has zero DC because of odd symmetry. The adjusted Gabor filter is used to filter the preprocessed images.

It should be pointed out that the success of 2D Gabor phase coding depends on the selection of Gabor filter parameters, θ, σ, and u. In our system, we applied a tuning process to optimize the selection of these three parameters. As a result, one Gabor filter with optimized parameters, $\theta=n/4$, $u=0.0916$, and $\sigma = 5.6179$ is exploited to generate a feature vector with 2,048 dimensions.

D. Hamming Distance

Given two data sets, a matching algorithm determines the degree of similarity between them. To describe the matching process cleary, we use feature matrices, real and imaginary. A normalized Hamming distance used in [6] is adopted to determine the similarity measurement for palmprint matching. Let P and Q be two palmprint and speech signal vectors. The normalized hamming distance can be described as

$$D_0 = \frac{\sum_{i=1}^{N}\sum_{j=1}^{N} P_M(i,j)\cap Q_M(i,j)\cap(P_R(i,j)\otimes Q_R(i,j)) + P_M(i,j)\cap Q_M(i,j)\cap(P_I(i,j)\otimes Q_I(i.j))}{2\sum_{i=1}^{N}\sum_{j=1}^{N} P_M(i,j)\cap Q_M(i,j)} \quad (8)$$

where $P_R(Q_R)$, $P_I(Q_I)$ and $P_M(Q_M)$ are the real part, the imaginary part and the mask of P(Q), respectively. The result of the Boolean operator (⊗) is equal to zero if and only if the two bits, $P_{R\ (i)}\ (i.j)$, are equal to $Q_{R\ (i)}\ (i.j)$; the symbol ∪ represents the AND operator and, the size of the feature matrixes is NxN. It is noted that D_0 is between 1 and 0. For the best matching, the hamming distance should be zero. Because of imperfect preprocessing, we need to vertically and horizontally translate one of the features and match again. The ranges of the vertical and horizontal translations are defined from -2 to 2. The minimum D_0 value obtained from the translated matching is considered to be the final matching score.

IV. FUSION

The biometrics systems is integrated at multi-modality level to improve the performance of the verification system. At multi-modality level, matching score are combined to give a final score. The following steps are performed for fusion:

1. Given a query image and speech signal as input, features are extracted by the individual recognition and then the matching score of each individual trait is calculated.
2. The weights a and b are calculated using FAR and FRR.
3. Finally, the final score after combining the matching score of each trait is calculated by weighted sum of score technique.

$$MS_{fusion} = \frac{a * MS_{Palm} + b * MS_{Speech}}{2} \quad (9)$$

Where a and b are the weights assigned to both the traits. The final matching score (MS_{fusion}) is compared against a certain threshold value to recognize the person as genuine or an imposter.

V. EXPERIMENTAL RESULTS

We evaluate the proposed multimodal system on a data set including 720 pairs of images from 120 subjects. The training database contains a speech signals and palmprint images for each individual for each subject. Each subject has 6 palm images taken at different time intervals and 6 different words, which is stored in the database. Before extracting features of palmprint, we locate palmprint images to 128x128.

The accuracy of Unimodal vs Multimodal is as shown in Fig. 5. The multimodal system has been designed at matching score level. At first experimental the individual systems were developed and tested for FAR, FRR & accuracy. In the last experiment both the traits are combined at matching score level using sum of score technique. The results are found to be very encouraging and promoting for the research in this field. The overall accuracy of the system is more than 97%, FAR & FRR of

2.8% & 0.8% respectively. Table1 shows FAR, FRR & Accuracy of the systems.

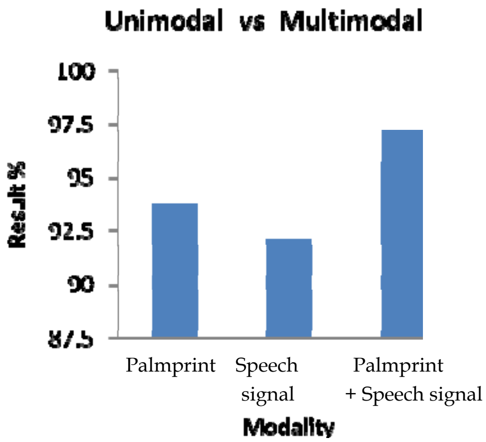

Fig. 5 Unimodal vs Multimodal

TABLE I
ACCURACY, FAR, FRR OF INDIVIDUAL RECOGNITION AND AFTER FUSION

Trait	Algorithm	FAR(%)	FRR(%)	Accuracy(%)
Palmprint	Gabor + Hamming distance	6.2	1.3	93.8
Speech Signal	MFCC + GMM	4.3	7.4	92.1
Palmprint + Speech Signal	weighted sum of score techniques	2.8	0.8	97.2

VI. CONCLUSION

Biometric systems are widely used to overcome the traditional methods of authentication. But the unimodal biometric system fails in case of biometric data for particular trait. Thus the individual score of two traits (speech signal & palmprint) are combined at classifier level and trait level to develop a multimodal biometric system. The performance table shows that multimodal system performs better as compared to unimodal biometrics with accuracy of more than 97%.

REFERENCES

[1] A. A. Ross, K. Nandakumar, and A. K. Jain. *Handbook of Multibiometrics*. Springer-Verlag, 2006..

[2] G. Feng, K. Dong, D. Hu & D. Zhang, *When Faces are Combined with Palmprints: A Noval Biometric Fusion Strategy,* ICBA, pp.701-707, 2004.

[3] G. Feng, K. Dong, D. Hu, and D. Zhang. When Faces Are Combined with Palmprints: A Novel Biometric Fusion Strategy. In *Proceedings of ICBA*, pages 701–707, 2004.

[4] G. Feng, K. Dong, D. Hu & D. Zhang, *When Faces are Combined with Palmprints: A Noval Biometric Fusion Strategy,* ICBA, pp.701-707, 2004.

[5] D. A. Reynolds, “Experimental Evaluation of Features for Robust Speaker Identification,” IEEE Transactions on SAP, vol. 2. Pp. 639-643,1994.

[6] J.G. Daugman, “*High Confidence Visual Recognition of Persons by a Test of Statistical Independence”, IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 15, no. 11, pp. 1148-1161, Nov. 1993.