Volume 1, No. 2, July-August 2010

ISSN No. 0976-5697

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

A Secured Multimodal Biometrics System using Palmprint & Speech Signal

Mahesh P.K.*
Research Scholar, JSS Research Foundation,
S.J.C.E., Mysore, Karnataka
India
mahesh24pk@gmail.com

M.N. ShanmukhaSwamy
Professor, JSS Research Foundation,
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 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 and palmprint recognition.

Keywords: Biometrics; multimodal; speech; 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 [8]. Speech and palmprint multimodal biometrics are advantageous due to the use of non-invasive and low-cost image acquisition. We can easily acquire speech signals and palmprint image using microphone and touchless sensor simultaneously. Existing studies in this approach [9, 10] employ holistic features for speech signal representation and results are shown with small data sets were reported. Note that holistic features are sensitive to global variation, such as illumination and inaccurate alignment.

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 [4] at the feature level fusion scheme. The proposed paper shows that integration of speech and palm print biometrics can achieve higher performance that may not be possible using a single biometric indicator alone. Both MFCC and 2D Gabor filter are considered in this feature vector fusion context.

The rest of this paper is organized as fallows. Section2 presents the system structure, which is used to increase the performance of individual biometric trait; multiple classifiers are combined using matching scores. Section 3 presents feature extraction using MFCC method and 2D Gabor. Section4, the individual traits are fused at matching score level using weighted sum of score techniques. Finally, the experimental results are given in section 5. Conclusions are given in the last section.

A. Combination of MFCC and 2DGabor Filter

The matching scores from the above two classifiers are converted from distance to similarity score and are combined at matching score level using sum of score technique which significantly increases the accuracy of the recognition system.

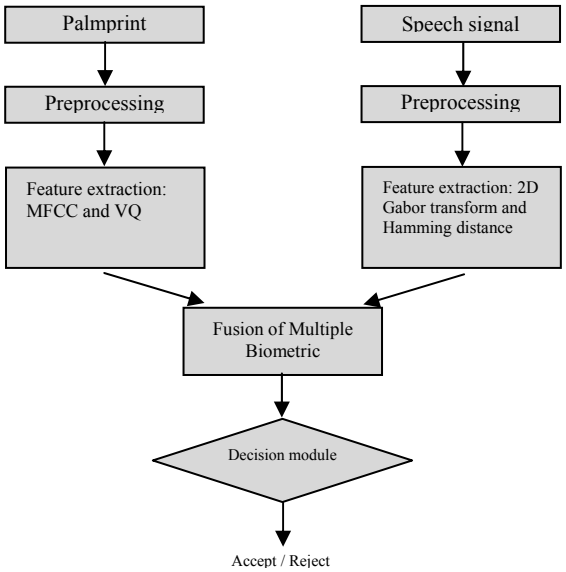

Figure 1. Block diagram of multimodal biometric system

© 2010, IJARCS All Rights Reserves

II. SYSTEM STRUCTURE

The multimodal biometric system is developed using two traits i.e. speech signal and palmprint as shown in Figure 1. For the speech signal and Palmprint Recognition, the input image is recognized using the combination feature of MFCC and 2D Gabor filter algorithm. When we are using a Gabor filter, the matching score is calculated using Hamming distance also when we are using MFCC; the matching score is calculated using VQ and Euclidean distance.

The modules based on the individual traits returns an integer valve after matching the database and query feature vectors. Here first the fusion is done at classifier level, i.e. for speech signal & palmprint, multiple classifiers are combined at matching score level followed by fusion at multiple modalities level. The final score is generated by using sum of score technique at matching score level, which is passed to the decision module.

III. FEATURE EXTRACTION USING MFCC FOR SPEECH SIGNAL

A. Speech Feature Extraction

The purpose of this module is to convert the speech waveform to some type of parametric representation (at a considerably lower information rate). The speech signal is a slowly time varying signal (it is called *quasi-stationar*y). When examined over a sufficiently short period of time (between 5 and 100 ms), its characteristics are fairly stationary. However, over long periods of time (on the order of 0.2s or more) the signal characteristics change to reflect the different speech sounds being spoken. Therefore, short-*time spectral analysis is* the most common way to characterize the speech signal. A wide range of possibilities exist for parametrically representing the speech signal for the speaker recognition task, such as Linear Prediction Coding (LPC), Mel-Frequency Cepstrum Coefficients (MFCC), and others. MFCC is perhaps the best known and most popular, and this feature has been used in this paper. MFCC's are based on the known variation of the human ear's critical bandwidths with frequency. The MFCC technique makes use of two types of filter, namely, linearly spaced filters and logarithmically spaced filters. To capture the phonetically important characteristics of speech, signal is expressed in the Mel frequency scale. This scale has a linear frequency spacing below 1000Hz and a logarithmic spacing above 1000 Hz. Normal speech waveform may vary from time to time depending on the physical condition of speakers' vocal cord. Rather than the speech waveforms themselves, MFFCs are less susceptible to the said variations.

B. The MFCC Processor

A block diagram of the structure of an MFCC processor is given in Figure 2. The speech input is recorded at a sampling rate of 22050Hz. This sampling frequency is chosen to minimize the effects of aliasing in the analog-to-digital conversion process. Fig. 2.shows the block diagram of an MFCC processor.

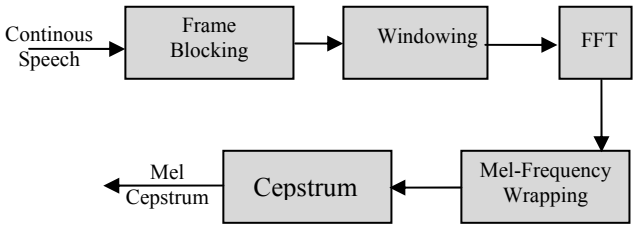

Figure 2. Block diagram of the MFCC processor

C. Mel-frequency wrapping

The speech signal consists of tones with different frequencies. For each tone with an actual Frequency, f, measured in Hz, a subjective pitch is measured on the 'Mel' scale. The mel-*frequency* scale is linear frequency spacing below 1000Hz and a logarithmic spacing above 1000Hz. As a reference point, the pitch of a 1kHz tone, 40dB above the perceptual hearing threshold, is defined as 1000 mels. Therefore we can use the following formula to compute the Mels for a given frequency *f* in Hz [5]:

$$\text{mel}(f) = 2595 * \log_{10}(1 + f/700) \quad (1)$$

One approach to simulating the subjective spectrum is to use a filter bank, one filter for each desired mel-frequency component. The filter bank has a triangular bandpass frequency response, and the spacing as well as the bandwidth is determined by a constant mel-frequency interval.

D. Cepstrum

In the final step, the log mel spectrum has to be converted back to time. The result is called the mel frequency cepstrum coefficients (MFCCs). The cepstral representation of the speech spectrum provides a good representation of the local spectral properties of the signal for the given frame analysis. Because the mel spectrum coefficients are real numbers (and so are their logarithms), they may be converted to the time domain using the Discrete Cosine Transform (DCT). The MFCCs may be calculated using this equation:

$$C\tilde{n} = \sum_{k=1}^{k} \left(\log \tilde{S}_k\right)[n(k - \tfrac{1}{2})\tfrac{\pi}{k} \quad where\ n = 1, 2.....K \quad (2)$$

The number of mel cepstrum coefficients, K, is typically chosen as 20. The first component, c_0, is excluded from the DCT since it represents the mean value of the input signal which carries little speaker specific information. By applying the procedure described above, for each speech frame of about 30 ms with overlap, a set of mel-frequency cepstrum coefficients is computed. This set of coefficients is called an *acoustic vecto*r. These acoustic vectors can be used to represent and recognize the voice characteristic of the speaker[4]. Therefore each input utterance is transformed into a sequence of acoustic vectors.

E. Vector Quantization

Vector quantization (VQ) is a lossy data compression method based on principle of block coding [6]. It is a fixed-to-

© 2010, IJARCS All Rights Reserves

fixed length algorithm. VQ may be thought as an aproximator. Figure 3 shows an example of a 2- dimensional VQ.

Here, every pair of numbers falling in a particular region are approximated by a star associated with that region. In Figure 3, the stars are called *code vectors* and the regions defined by the borders are called encoding *regions*. The set of all code vectors is called the *codebook* and the set of all encoding regions is called the *partition* of the space [6].

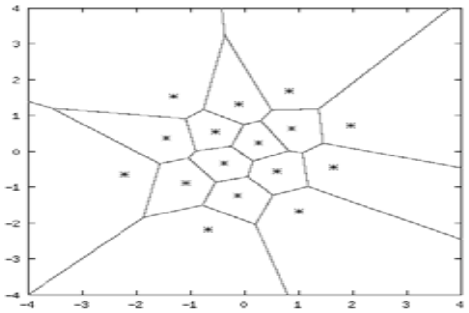

Figure 3. An example of a 2-dimensional VQ

IV. FEATURE EXTRACTION USING GABOR TRANSFORM FOR PALMPRINT

A. Feature Extraction and Coding (Gabor Filter)

We proposed a 2D Gabor phase coding scheme for palmprint recognition [9] . The circular Gabor filter is an effective tool for texture analysis [12], and has the following general form.

$$G(x,y,\theta,u,\sigma,) = \frac{1}{2\pi\sigma^2}\exp- \left\{\frac{x^2-y^2}{2\sigma^2}\right\}\exp\{2ni(ux\cos\theta + uy\sin\theta)\} \quad (3)$$

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:

$$\tilde{G}[s,y,\theta,u,\sigma] = G[x,y,\theta,u,\sigma]\frac{-\sum_{i=-n}^{n}\sum_{j=-n}^{n}G[i,y,\theta,u,\sigma]}{(2n+1)^2} \quad (4)$$

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, *θ=n/4, u=0.0916,* and *σ = 5.6179* is exploited to generate a feature vector with 2,048 dimensions.

B. Euclidean Distance

Let an arbitrary instance X be described by the feature vector $X=[a_1(x),a_2(x),\ldots a_n(x)]$ where $a_r(x)$ denotes the value of the r^{th} attribute of instance x. Then the distance between two instances x_i and x_j is defined to be $d(x_i,x_j)$;

$$d(x_i, x_j) = \sqrt{\sum_{r=1}^{n}(a_r(X_i) - a_r(X_j))^2} \quad (5)$$

V. FUSION

Score combination method uses operation to combine the individual scores (as can be the weighted sum or the weighted product), and the result is compared with some decision threshold. Here we use the weighted sum method. Here the combined score is weighted sum of unimodal scores. The decision is also calculated by comparing this score with a threshold. It is a little bit more expensive because it's necessary to implement the power (pow) function.

The combined opinions from speech and palm images using the weighted sum approach:

$$f= (o_1)^{\alpha} + (o_2)^{1-\beta} \quad (6)$$

Where o_1 and o_2 are the opinions from the speech signal and palm profile experts, respectively, with corresponding weights α and β. Each opinion reflects the likelihood that a given claimant is a true claimant (i.e., a low opinion suggests that the claimant is an impostor, while high opinion suggests that the claimant is the true claimant).

$$\beta = \frac{1-(FAR_2 + FRR_2)}{2-(FAR_1 + FRR_1 + FAR_2 + FRR_2)} \quad (7)$$

$$\alpha = \frac{1-(FAR_1 + FRR_1)}{2-(FAR_2 + FRR_2 + FAR_1 + FRR_1)} \quad (8)$$

Where FAR_1 and FRR_1 are the False Acceptance Rate and False Rejection Rate of speech signal and FAR_2 and FRR_2 are False Acceptance Rate and False Rejection Rate of palmprint.

VI. 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.

Whe

© 2010, IJARCS All Rights Reserves

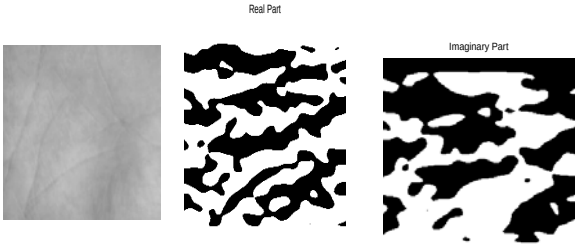

Figure 4. Gabor based Palm images (a) Preprocessed Image, (b) Real Part of Texture Image, and (c) Imaginary Part of Texture Image

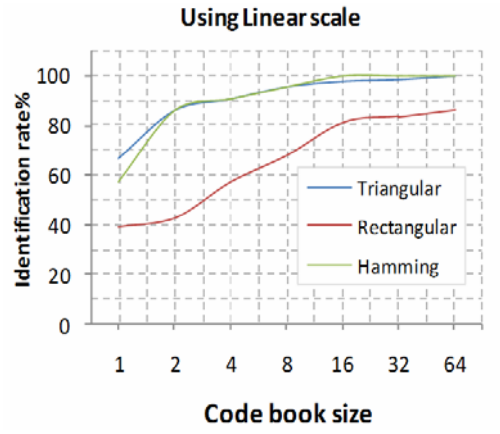

*Figure 5. Iendtification rate (in %) for different windows [using **Linear scale**]*

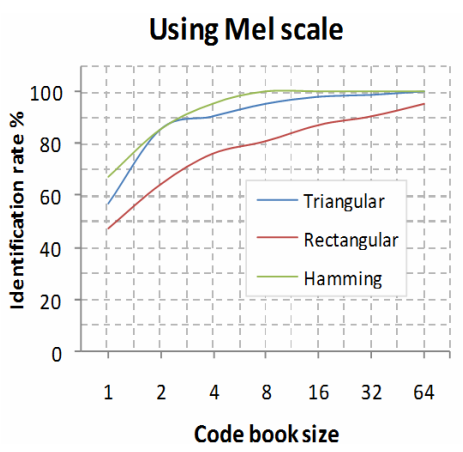

*Figure 6. Identification rate (in %) for different windows [using **Mel scale**]*

Table I. FAR, FRR & Accuracy of the individual system and after fusion

Trait	Algorithm	FAR	FRR	Accuracy
Speech signal	MFCC + VQ	6.59%	4.76%	92.8%
Palmprint	Gabor Transform + Hamming Distance	5.49%	1.2%	94%
Speech Signal + Palmprint	sum of score technique.	2.6%	0.8%	97.2%

Figure 5 shows identification rate when triangular, or rectangular or hamming window is used for framing in a linear frequency scale. The table clearly shows that as codebook size increases, the identification rate for each of the three cases increases, and when codebook size is 16, identification rate is 100% for the hamming window. However, in case of Fig. 6 the same windows are used along with a Mel scale instead of a linear scale. Here, too, identification rate increases with increase in the size of the codebook. In this case, 100% identification rate is obtained with a codebook size of 8 when hamming window is used.

Table I shows FAR, FRR & Accuracy of the individual system and after fusion. 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 1.67% & 0.8% respectively.

VII. 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 matching score 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%.

VIII. References

[1] Lawrence Rabiner and Biing-Hwang Juang, Fundamental of Speech Recognition", Prentice-Hall, Englewood Cliffs, N.J., 1993.

[2] Zhong-Xuan, Yuan & Bo-Ling, Xu & Chong-Zhi, Yu. (1999). "Binary Quantization of Feature Vectors for Robust Text-Independent Speaker Identification" in *IEEE Transactions on Speech and Audio Processing, Vol. 7, No. 1*, January 1999. IEEE, New York, NY, U.S.A.

[3] F. Soong, E. Rosenberg, B. Juang, and L. Rabiner, "A Vector Quantization Approach to Speaker Recognition", AT&T Technical Journal, vol. 66, March/April 1987, pp. 14-26

[4] Comp.speech Frequently Asked Questions WWW site, http://svr-www.eng.cam.ac.uk/comp.speech/

[5] Jr., J. D., Hansen, J., and Proakis, J. Discrete Time Processing of Speech Signals, second ed. IEEE Press, New York, 2000.

[6] R. M. Gray, "Vector Quantization",' IEEE ASSP Magazine, pp. 4--29, April 1984.

[7] Y. Linde, A. Buzo & R. Gray, "An algorithm for vector quantizer design", IEEE *Transactions on Communication*s, Vol. 28, pp.84-95, 1980..

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

[9] A. Kumar and D. Zhang. Integrating palmprint with face for user authentication. In *Proc.Multi Modal User Authentication Workshop*, pages 107–112, 2003.

[10] 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.

© 2010, IJARCS All Rights Reserves