Score Level Fusion Based Multimodal Biometric Identification (*Fingerprint & Voice*)

Youssef ELMIR
dept. of computer science
African University of Ahmed Draïa,
Adrar, Algeria
ysfbr@yahoo.fr

Zakaria ELBERRICHI, Reda ADJOUDJ
dept. of computer science
Djillali Liabes University, UDL
Sidi bel Abbes, Algeria
elberrichi@gmail.com, adjreda@yahoo.fr

***Abstract*— Feature level based monomodal biometric systems perform person recognition based on a multiple sources of biometric information and are affected by problems like integration of evidence obtained from multiple cues and normalization of features codes since they are heterogeneous, in addition of monomodal biometric systems problems like noisy sensor data, non-universality and lack of individuality of the chosen biometric trait, absence of an invariant representation for the biometric trait and susceptibility to circumvention. Some of these problems can be alleviated by using multimodal biometric systems that consolidate evidence from scores of multiple biometric systems. In this work, we address two important issues related to score level fusion. We have studied the performance of a score level fusion based multimodal biometric system against different monomodal biometric system based on voice, fingerprint modalities and a bimodal biometric system based on feature level fusion of the same modalities. These systems have been evaluated in terms of their efficiency and identification rate on a close group from the test data. These results are shown using cumulative match characteristic curve.**

Keywords-component; Fusion strategies; Multimodal biometrics; SVM; Gabor filters; MFCC.

I. Introduction

Automatic extraction of identity cues from personal traits (e.g., fingerprints, speech, and face images) has given rise to a particular branch of pattern recognition (biometrics) where the goal is to infer identity of people from biometric data [1]. Our efforts have been focused on two basic biometric characteristics, namely, on-line fingerprint and voice which are physiological characteristics, due to the following reasons: i) regarding fingerprint, due to its uniqueness and high discriminative capability; ii) regarding voice, for its personal, social and legal acceptability as an identification procedure.

In this contribution, and after reviewing some referenced approaches to fusion in multimodal biometrics, especially those based on SVM classifiers as they have shown outstanding performance [2] we will derive fusion schemes based on the generation of a combined score. Thereafter, recognition rates will be compared using our Fingerprint, voice and feature fusion based systems on FingerCell and ELSDSR databases.

II. Creating Codes

A. *Fingerprint code*

Following steps are observed to create the fingercode:

a) Preprocessing of the image (to remove noise) by window wise normalization, Histogram Equalization, low pass and median filtering [3].

b) Core point location using max concavity estimation [4].

c) Tessellation of circular region around the reference point.

d) Sector wise normalization followed by application of bank of Gabor filters which has a general form (see (1)) in the spatial domain [5].

$$G(x,y;f,\theta)=\exp\left\{\frac{-1}{2}\left[\frac{x'^2}{\delta_{x'}^2}+\frac{y'^2}{\delta_{y'}^2}\right]\right\}\cos(2\pi f x') \quad (1)$$

With $x' = x\sin\theta + y\cos\theta$ and $y' = x\sin\theta - y\cos\theta$.

Where f is the frequency of the sine plane wave along the direction θ from the x-axis, and $\delta_{x'}$ and $\delta_{y'}$ are the space constants of the Gaussian envelope along X’ and Y’ axes, respectively.

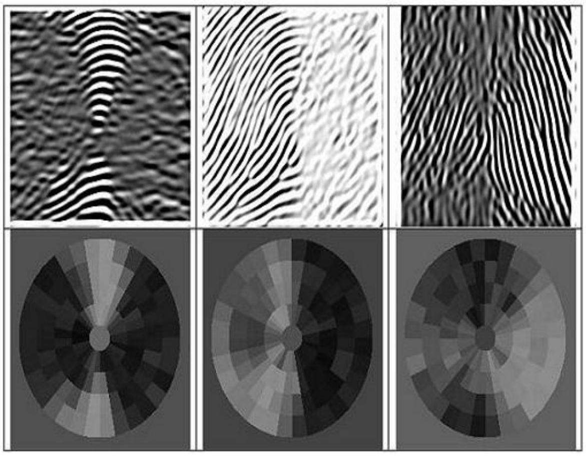

Figure 1. Filtered images and their corresponding feature vectors for orientations 0°, 5°. 22.5° and 45° are shown [6]

5. Finally feature code generation by obtaining standard deviation values of all the sectors, [5].

B. Voice code

The signal of a voice is first processed by software that convert the speech waveform to some type of parametric representation (at a considerably lower information rate) for further analysis and processing. The speech signal is a slowly timed varying signal (it is called quasi-stationary). An example of speech signal is shown in fig. 2. When examined over a sufficiently short period of time (between 5 and 100 msec), its characteristics are fairly stationary. However, over long periods of time (on the order of 1/5 seconds or more) the signal characteristic 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), Gaussian mixture models (GMM) [15], Mel-Frequency Cepstrum Coefficients (MFCC), and others.

MFCC are perhaps the best known and most popular, and these will be used in this paper. MFCC's are based on the known variation of the human ear's critical bandwidths with frequency, filters spaced linearly at low frequencies and logarithmically at high frequencies have been used to capture the phonetically important characteristics of speech(to obtain voiceprint or voice signal matrix code). This is expressed in the mel-frequency scale, which is a linear frequency spacing below 1000 Hz and a logarithmic spacing above 1000 Hz. The process of computing MFCCs is described in more detail in [7], [8].

A block diagram of the structure of an MFCC processor is given in fig. 2. The speech input is typically recorded at a sampling rate above 10000 Hz. This sampling frequency was chosen to minimize the effects of aliasing in the analog-to-digital conversion. These sampled signals can capture all frequencies up to 5 kHz, which cover most energy of sounds that are generated by humans. As been discussed previously, the main purpose of the MFCC processor is to mimic the behavior of the human ears. In addition, rather than the speech waveforms themselves, MFCC's are shown to be less susceptible to mentioned variations. The voice signal matrix is immediately encrypted to eliminate the possibility of identity theft and to maximize security. For example, here is the voice signal from database [8] and the voiceprint matrix of this voice signal (Fig. 2).

Perfect classification of N ideal input voiceprint matrix of voice signal is required, and reasonably accurate classification of speech waveform (N is equivalent to a number of distinguished class of speaker in each database). The N 160-element input voiceprint matrix of voice signals are defined as a matrix of input matrixes (voiceprint matrix size ~ 20 x 8). The target vectors are also defined with a variable called targets. Each target vector is a N-element vector with a 1 in the position of the voiceprint it represents, and 0's everywhere else. For example, the voiceprint number one is to be represented by a 1 in the first element (as this example is the first voiceprint of the database), and 0's in elements two through N.

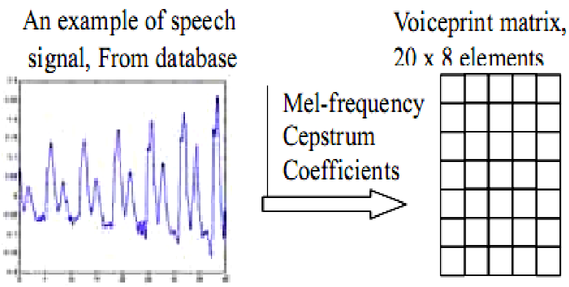

Figure 2. Voiceprint

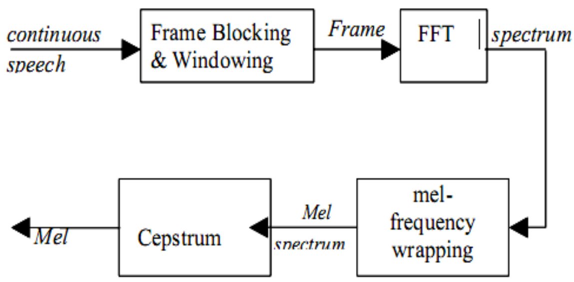

Figure 3. Block diagram of the MFCC processor

III. MULTIMODAL FUSION

A. *Fusion strategies*

Biometric multimodality can be studied as a classifier combination problem [9], [10]. Kittler et al. considered in [10] the task of combining classifiers in a probabilistic bayesian framework and provided an example of multimodal biometric verification (fusing speech, frontal and profile images modalities). Several ways to merge the modalities are obtained (sum, product, max, min, …), based on the Bayes theorem and certain hypothesis, from which the Sum Rule (i.e., the combined score is obtained adding the monomodal scores which have been previously mapped to the [0,1] range) outperformed the remainder in the experimental comparison due to its robustness to errors made by the individual classifiers. From now on, this perspective will be referred to as rule-based fusion, because it does not takes into account the actual distribution of outputs from the experts.

Multimodal fusion can also be treated as a pattern classification problem [11]. Under this point of view, the scores given by individual expert modalities are considered as input patterns to be labeled as accepted/rejected (for the verification task). Verlinde et al. followed this approach and compared in [12] the following pattern classification techniques for multimodal fusion (sorted by relative decreasing performance): Logistic Regression, Maximum a Posteriori, k-Nearest Neighbours classifiers, Multilayer Perceptrons, Binary Decision Trees, Maximum Likelihood, Quadratic classifiers and Linear classifiers. In a recent contribution [13], the paradigm of Support Vector Machines (SVMs) has been compared with all the above-mentioned techniques carrying out the same experiments, outperforming all of them. From now on, this perspective will be referred to as learning-based (or trained) fusion, because it requires sample outputs from the experts to train the pattern classifiers.

In our case, we have chosen score based fusion strategy, using score obtained from a fingerprint identification system and the one obtained from the speaker identification system.

B. *Multimodal Fusion*

We have used the SVM in order to provide not a binary verification decision, as it has been reported in related works [14][9], but rather a merged score combining the outputs of the considered monomodal experts. We will now introduce our approach providing references for further details.

The principle of SVM relies on a linear separation in a high dimension feature space where the data have been previously mapped, in order to take into account the eventual non-linearities of the problem [15]. In order to achieve a good level of generalization capability, the margin between the separator hyperplane and the data is maximized.

IV. EXPERIMENTS

A. *Databases Description*

We have selected first 10 users from the FingerCell fingerprints Database and the same number of users from the ELSDSR voices database [16], and thanks to the independence of fingerprint and voice traits [13], we have created 10 individuals comprising fingerprint and voice traits.

The following training and testing procedure for monomodal systems had been established.

- Training:

  i) Fingerprint: Each client’s index finger has been represented with 1 high-control minutiae pattern;

  ii) Voice: Each voice signal has been modeled with one sample.

- Testing:

  a) One sample of each trait (fingerprint and voice) has also been selected for tests;

B. *Monomodal Fusion*

Standard performance individual verification systems (whose parameters have not been optimized) have been intentionally used because it makes the comparison of subsequent fusion strategies easier. In particular, we have considered: a fingerprint recognition system based on gabor filters [17], a Mel-frequency cepstral coefficients-based voice recognition system. For the monomodal based approach, a unique SVM is used for all users and the leave-one-out method [11], leaving out each one of the users, will be applied for testing (i.e., feature code of each user will be trained with SVM).

C. *Feature Fusion based Multimodal System*

For the feature fusion level based approach, all multimodal combined test feature codes (fingerprint and voice) are used for testing the recognition performance of the trained SVM.

D. *Score Fusion based Multimodal System*

For the score fusion level based approach, all monomodal combined test scores obtained from each SVM are used for testing the recognition performance.

V. EXPERIMENTS

TABLE I. PERCENTAGE OF RECOGNITION USING PC WITH 2.4 GHZ CORE 2 DUO PROCESSOR

Biometric recognition system	***Recognition Rate (%)***
Speaker identification	50
Fingerprint Identification	60
Feature level fusion	70
Score level fusion	70

Table 1 shows that the biometric systems based on fusion gave better recognition rate than monomodal systems. However, both of score and feature level based identification systems achieved same identification rate.

In order to make a good comparison between these two fusion based system, we have made multiple test regarding the rank of identification. This has given clear results which are shown in fig. 4.

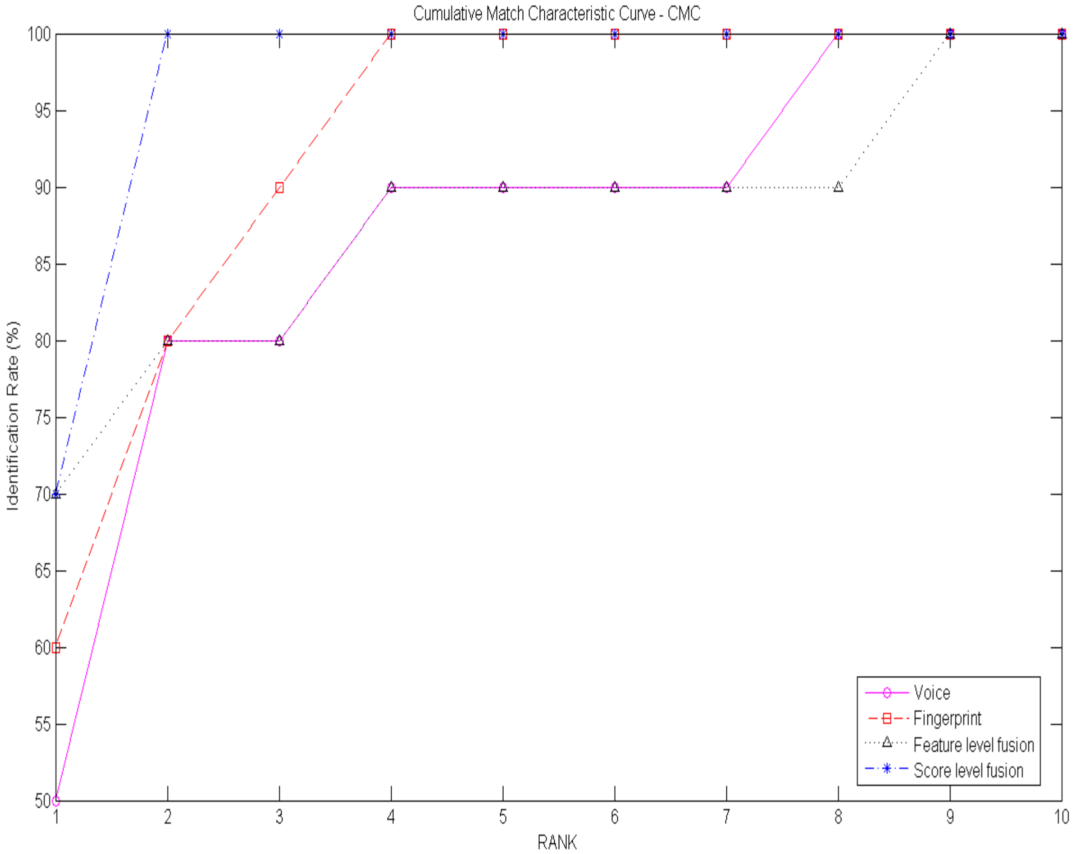

Figure 4. *Performance of Monomodal and Multimodal systems*

In fig. 4, the performance of monomodal and multimodal systems are plotted together. It is shown that score level fusion based system gave the better identification rate (100%) since the second rank, however, feature level fusion based system waited until the rank 9 to achieve 100%. In the other side, monomodal systems fingerprint and voice still far from this high rate until last ranks 4 and 8 respectively.

VI. CONCLUSION

A statistical-motivated experimental procedure has been introduced and applied in order to compare best referenced fusion based and monomodal based biometric recognition rate by means of DET plots. SVM classifier has been proposed to classify fusion codes have been derived

Appropriate selection of parameters for the score fusion based approach has shown to provide better recognition performance than the monomodal based approach. In particular, starting from, approximately, a 50% error rate speaker identification system, a 40% error rate fingerprint identification system and a 30% error rate multimodal identification system, it has been shown that the score fusion based multimodal system reduced the error rate to 0% in the second rank of test. Encouraging initial results of this approach motivate further research in order to exploit user specificities in the fusion stage of multimodal biometric recognition systems on large data sets.

REFERENCE

[1] A. Jain, R. Bolle and S. Pankanti (eds.), *Biometrics – Personal Identification in Networkwed Society*, Kluwer Academic Publishers, 1999.

[2] L. Xu, A. Kryzak and C.Y. Suen, "Methods of Combining Multiple Classifiers and Their Application to Handwriting Recognition", *IEEE Trans. on Systems, Man and Cybernetics*, vol. 22, no. 3, pp. 418-435, May-June 1992.

[3] S Greenberg, M Aladjem, D Kogan and I Dimitrov, "Fingerprint Image Enhancement using Filtering Techniques" *Proc of the 15th International Conference on Pattern Recognition (ICPR'00)*, Barcelona, Spain, September 2000

[4] A. K. Jain, D Maltoni, D Maio, S Prabhakar, "*Handbook of Fingerprint Recognition*", Springer Professional Computing, 2003.

[5] A K. Jain, S Prabhakar, and L Hong, "A Multichannel Approach to Fingerprint Classification", in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 4, pp. 348–359, April 1999.

[6] M U Munir and M Y Javed, College of Electrical and Mechanical Engineering, National University of Sciences and Technology Rawalpindi, Pakistan, 2004.

[7] Minh N. Do, "An Automatic Speaker Recognition System", Audio Visual Communications Laboratory, Federal Institute of Technology, Lausanne, Switzerland, February, 2003.

[8] C. Cornaz, U. Hunkeler an V.Velisavljevic, "An automatic speaker recognition system", Digital Signal Processing Laboratory, Federal Institute of Technology, Lausanne, Switzerland, February, 2003.

[9] B. Gutschoven and P. Verlinde, "Multi-Modal Identity Verification using Support Vector Machines (SVM)", *Proc. of the 3rd Intl. Conf. on Information Fusion*, 2000.

[10] J. Kittler, M. Hatef, R.P.W. Duin and J. Matas, "On Combining Classifiers", *IEEE Trans. Pattern Anal. and Machine Intell.*, vol. 20, no. 3, pp. 226-239, March 1998.

[11] S. Theodoridis and K. Koutroumbas, "*Pattern Recognition*", *Academic Press, 1999.*

[12] P. Verlinde, G. Chollet and M. Acheroy, "Multi-Modal Identity Verification using Expert Fusion", Information Fusion, no. 1, pp. 17-33, *Elsevier*, 2000.

[13] A. Ross, A. K. Jain, J. Z. Qian, "Information Fusion in Biometrics", *Proc. of the 3rd Audio and Video-Based Person Authentication*, AVBPA, pp. 354-359, Halmstad, Sweden, 2001.

[14] M. Young, "*The Technical Writer's Handbook".* Mill Valley, CA, University Science, 1989.

[15] Y. Mami, "Reconnaissance de locuteurs par localisation dans espace de locuteurs de références", PhD thesis, at « Ecole nationale supérieure des télécommunications », Paris, France, October 21,2003.

[16] Feng, L., Speaker Recognition, Informatics and Mathematical Modelling, Technical University of Denmark, DTU, 2004.

[17] Y. Elmir and M. Benyettou, "Gabor Filters Based Fingerprint Identification Using Spike Neural Networks", *Proc of the 5th Sciences of Electronic, Technologies of Information and Telecommunication conference*, Hammamet, Tunisia, March 2009.