Impact Factor (SJIF): 4.542

International Journal of Advance Research in Engineering, Science & Technology

e-ISSN: 2393-9877, p-ISSN: 2394-2444

Volume 4, Issue 5, May-2017

A Review of Techniques of Fusion of Speech and Image Recognition for Multimodal Biometric System

Ruchi J Misra[1], Komal R Borisagar[2], Pratik Kadechha[3]

[1]*Student,ME(EC)t , Atmiya Institute of Technology and Sciences*
[2]*Associate Professor(EC), Atmiya Institute of Technology and Sciences*
[3]*Lecturer(EC), Atmiya Institute of Technology and Sciences*

Abstract —***In Multimodal Biometrics, biometric traits are used together at a specific level of fusion to recognize persons. Either multiple algorithms called classifiers are used at enrolment or matching stages for various biometric traits or multiple sensors of the same biometric trait or multiple instances of the same biometric trait or repeated instances of the same biometric trait. Paper reviews use of different biometric traits like face, speech, fingerprint, iris, and various fusion scenarios.***

Keywords-fusion; biometric; multimodal; image recognition; speech recognition

I. INTRODUCTION

Conventional means of identification such as passwords, secret codes can easily be compromised, shared, observed, stolen or forgotten. However, a possible alternative in determining the identities of users is to use biometrics. Biometrics used for person recognition refers to the process of automatically recognizing a person distinguishing one of the qualities namely Behavioural patterns (gait, signature, keyboard typing, lip movement, hand-grip) or Physiological traits (face, voice, iris, fingerprint, hand geometry, electroencephalogram -- EEG, electrocardiogram -- ECG, ear shape, body odour, body salinity, vascular).

Several of these biometric modalities have been investigated (fingerprint, iris, voice, face) and are still under consideration. Face and voice biometrics have lower performance compared to other biometric traits. However, these constitute some of the most widely accepted by people, and the low cost of the equipment for face and voice acquisition makes the systems inexpensive to build. Their combinations have resulted in varying degrees of performances. More recently, novel biometric modalities have emerged (gait, EEG, vascular) mainly due to the development of sensor technologies.

Since biometric person recognition is a truly inter-disciplinary research field and offers a wide range of challenging problems in image processing, computer vision, pattern recognition and machine learning.

II. MULTIMODAL BIOMETRIC SYSTEM

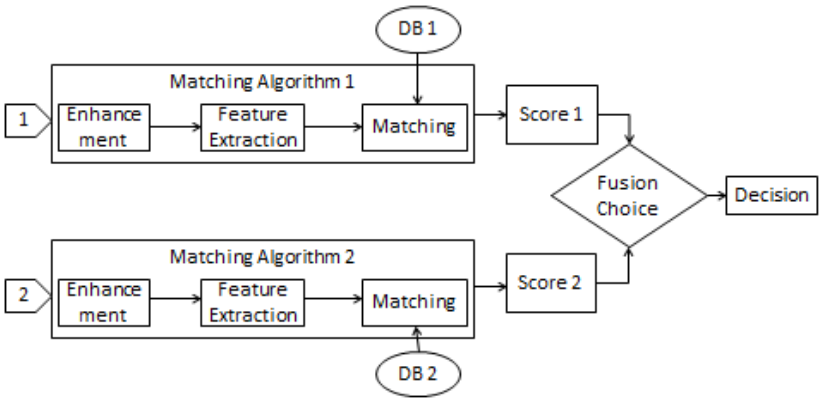

Figure 1: A Multimodal Biometric System(Score Level Fusion)

Recognition, identification, authentication and verification have different meanings. Recognition refers to the research field (Biometric person recognition) and is categorized in two modes: authentication (also called verification) and identification. An authentication (or verification) system involves confirming or denying the identity claimed by a person (one-to-one matching). In contrast, an identification system attempts to establish the identity of a given person out of a

closed pool of N people (one-to-N matching). Authentication and identification share the same preprocessing and feature extraction steps and a large part of the classifier design. However, both modes target distinct applications. In authentication mode, people are supposed to cooperate with the system (the claimant wants to be accepted).

2.1 Face Recognition

The challenge in face processing (detection and recognition) is that faces highly vary in size, shape, colour, texture and location. Their overall appearance can also be influenced by lighting conditions, facial expression, occlusion or facial features, such as beards, moustaches and glasses. Another challenging problem comes from the orientation (upright, rotated) and the pose (frontal to profile) of the face. Face detection is to determine whether or not there are any faces in the image and, if present, their location. It is the first step of any application that involves face processing systems. Thus, accurate and fast human face detection is the key to a successful operation. Face recognition has been an active research area for many years and different systems are now capable of correctly recognizing people's faces under specific environments (near frontal faces and controlled imaging conditions). However, many applications need the ability to deal with faces of varying head poses and adverse imaging conditions since most faces in the real world are not frontal and captured in uncontrolled environments[7].

2.2 Fingerprint Recognition

The fingerprint recognition system has been developed by the Minutiae Matching Techniques. The key steps involved are fingerprint enhancement, feature extraction using Minutiae Matching approach and computation of matching score. The goal of fingerprint enhancement is to increase the clarity of ridge structure so that minutiae and the reference points can be easily and correctly extracted[1][7].

2.3 Iris Recognition

The minute architecture of the iris exhibits variations in every person. A template created by imaging an iris is compared to stored template(s) in a database. If the Hamming distance is below the decision threshold, a positive identification has effectively been made because of the statistical extreme improbability that two different persons could agree by chance ("collide") in so many bits, given the high entropy of iris templates[2][1].

2.4 Speaker recognition

A speaker recognition system uses a speech utterance to determine if it has been pronounced by a known person. This is also a difficult task depending on the quality of the capture device, the conditions and of the cooperation of the subject. Generally, the first task is to extract the relevant information (speech frames) and to filter out irrelevant information (silence, ambient noise, music or background speech) before the actual speaker recognition is triggered.

Different scenarios can take place, namely text independent and text dependent. In text independent speaker recognition, the identity models are assumed to be independent of the precise sentence pronounced by a person. They use Gaussian Mixture Models(GMM)[13] for modelling and Log Likelihood Ratio(LLR) for comparing. In text dependent speaker recognition, the lexical content of the sentence pronounced by a person is more important and enables better robustness against replay attacks. However, text dependent speaker recognition system generally needs more resources than text independent ones to efficiently process this lexical information The algorithms used are comparison Mel Frequency Cepstral Coefficients[14], Hidden Markov Models (HMM) or Artificial Neural Networks(ANN).

III. LEVELS OF FUSION

In the past ten years, it has been shown that combining biometric systems achieves better performance than techniques using only one biometric modality. This has been shown to be true using various fusion algorithms. Fusion algorithms are methods whose goal is to merge the prediction of many algorithms (multiple biometric modules) in the hope of a better average performance than any of the individual methods. This fusion can be simple (maximum score, product or sum rules), but it is often better to train a fusion system using Machine Learning algorithms.

Most of the proposed fusion techniques, often called late integration techniques, operate at the score or decision levels. Other techniques, called early integration techniques, aim to exploit the correlation between biometric modalities if any. This is the case for instance between the video and audio streams of a talking face while the person pronounces a sentence like footage from a video camera[6][3].

3.1Fusion Levels

Fusion can be carried out at various levels[1]

3.1.1 Sensor Level. Multisensorial biometric systems sample the same instance of a biometric trait with two or more distinctly different sensors. Processing of the multiple samples can be done with one algorithm or combination of

algorithms. e.g. face recognition application could use both a visible light camera and an infrared camera coupled with specific frequency.

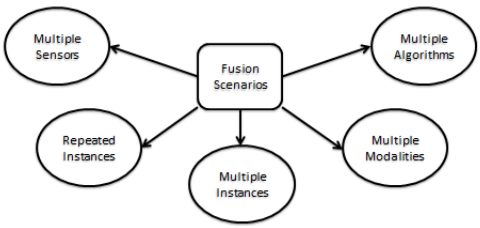

Figure 2. Scenarios of Fusion

3.1.2 Feature Level. The feature level fusion is useful in classification. Different feature vectors are combined, obtained either with different sensors or by applying different feature extraction algorithms to the same raw data.
3.1.3 Decision Level. With this approach, each biometric subsystem completes autonomously the processes of feature extraction, matching, and recognition. Decision strategies are usually of Boolean functions, where the recognition yields the majority decision among all present subsystems.
3.1.4 Rank Level. Instead of using the entire template, partitions of the template are used. Ranks from template partitions are consolidated to estimate the fusion rank for the classification. Rank level fusion involves combining identification ranks obtained from multiple unimodal biometrics. It consolidates a rank that is used for making final decision
3.1.5 Score Level. It refers to the combination of matching scores provided by the different systems.

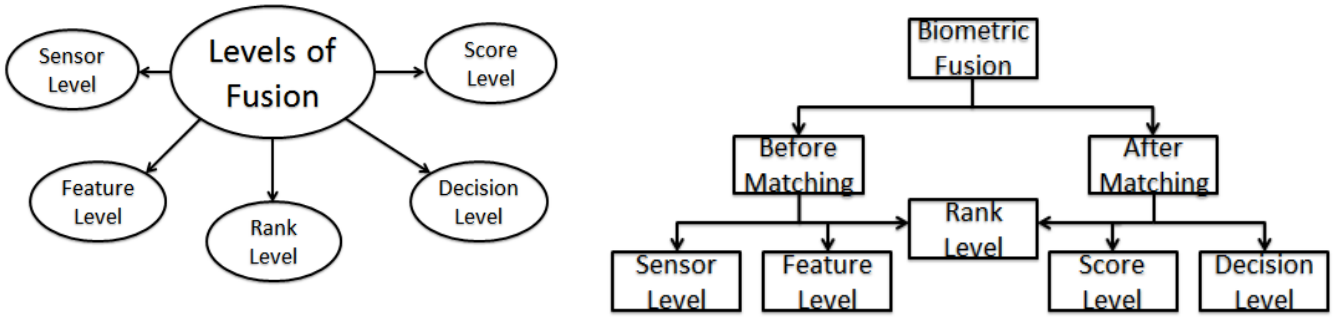

Figure 3. Levels of Fusion and their hierarchy

3.2. Score Rules.
The score level fusion techniques are divided into two main sets. The Fixed Rules like AND, OR, majority, maximum, minimum, sum, product and arithmetic rules[12]. And the Trained Rules like weighted sum, weighted product, fisher linear discriminate, quadratic discriminate, logistic regression, support vector machine, multilayer perceptrons, and Bayesian classifier. The Score Normalization Algorithms[12] are Min-Max (MM) normalization, Z-Score (ZS) normalization, Median and Median Absolute Deviation (MAD) normalization, Tanh normalization, Double-Sigmoid (DS) normalization, Modify Double-Sigmoid (MDS) normalization. The rules for score fusion Techniques are
3.2.1 Simple Sum Rule(SSR). Here the fused score is calculated by adding the scores of all the modalities involved. Mathematically

$$f = \sum_{m=1}^{M} x_m$$

3.2.2 Maximum Rule(MAR). Here score having largest value among the modalities involved is selected. Mathematically[12]

$$f = \max(x_1, x_2, \ldots, x_M)$$

3.2.3 Minimum Rule(MIR). Here match score represents the difference score. Minimum rule method score chooses the score having the least value of the modalities involved, and is defined as[12]

$$f = \min(x_1, x_2, \ldots, x_M)$$

3.2.4 Product Rule. Here fused score is calculated by multiplying the scores for all modalities involved and is mathematically defined as [12]

$$f = \prod_{m=1}^{M} x_m$$

3.3 Matching Strategies

The Classical Matching Strategies are Hamming Distance Based, Sum Rule Based or Weighted Sum Rule Based. On the other hand the are Fuzzy Logic Strategies[1] like If-then rules.

3.4 Fusion Techniques

The Late Integration Techniques operate at the score or decision levels. And the Early Integration Techniques aim to exploit the correlation between biometric modalities if any. e.g. between the video and audio streams of a talking face while the person pronounces a sentence.

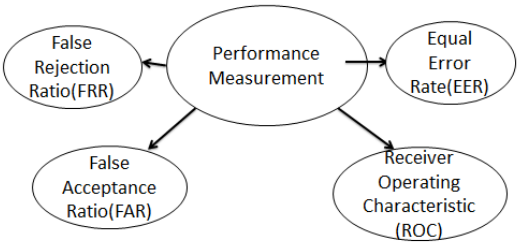

Figure 4 Performance Measurement

Performance

Like for any recognition System, performance is measured in terms of False Rejection Ratio, False Acceptance Ratio, Equal Error Rate and Receiver Operating Characteristics. ROC curve plots probability of FAR versus FRR.GAR(Genuine Acceptance Ratio) is also used as a parameter to compare performances. GAR(%) and ROC curves show that fusion of multiple biometrics improves the performance compared to single biometrics[2][7][9]. In addition to this performance is also compared in terms of time taken.

IV. SUMMARY

The main applications of access control systems (airport checking, monitoring, computer or mobile devices log-in), building gate control, digital multimedia access, transaction authentication (in telephone banking or remote credit card purchases for instance), voice mail, or secure teleworking. On the other hand, in identification mode, people are generally not concerned by the system and often even do not want to be identified. Potential applications includes video surveillance (public places, restricted areas), forensic (police databases) and information retrieval (video or photo album annotation/identification).

For image(here face, fingerprint and iris) recognition algorithms face recognition is cheaper to implement and easier to obtain but fingerprint and iris are more precise and need consent of user. The accuracy level of speech surpasses that of fingerprints[11]. A customized selection of traits incorporated in recognition and then utilized to implement decision should give better results. The various fusion algorithms and comparisons of the same in combination with various biometric traits need to be studied. The performance evaluation needs be done based on the results with respect to precision requirement and cost and computational expenses.

REFERENCES

[1] Houda Benaliouche and Mohamed Touahria, “Comparative Study of Multimodal Biometric Recognition by Fusion of Iris and Fingerprint”, in Hindawi Publishing Corporation, Scientific World Journal, Volume 2014.

[2] Sheetal Chaudhary and Rajender Nath, "A New Multimodal Biometric Recognition System Integrating Iris, Face and Voice", in International Journal of Advanced Research in Computer Science and Software Engineering,Volume 5, Issue 4, April 2015

[3] Dhaval Shah, Kyu J. Han and Shrikanth S. Nayaranan,"A Low-Complexity Dynamic Face-Voice Feature Fusion Approach to Multimodal Person Recognition", in ISM, 2009, 2013 IEEE International Symposium on Multimedia, 2013 IEEE International Symposium on Multimedia

[4] Miguel Carrasco, Luis Pizarro, and Domingo Mery,"Bimodal Biometric Person Identification System Under Perturbations", in PSIVT 2007, LNCS 4872, pp. 114–127, 2007.

[5] Muhammad Imran Razzak, Muhammad Khurram Khan,Khaled Alghathbar and Rubiyah Yusof, "MULTIMODAL BIOMETRIC RECOGNITION BASED ON FUSION OF LOW RESOLUTION FACE AND FINGER VEINS",in International Journal of Innovative Computing, Information and Control Volume 7, Number 8, August 2011

[6] Imran Naseem1 and Ajmal Mian2Bebis et al., "User Verification by Combining Speech and Face Biometrics in Video", in ISVC 2008, Part II, LNCS 5359, Springer-Verlag Berlin Heidelberg 2008

[7] Anil Jain, Lin Hong, Yatin Kulkarni, A Multimodal Biometric System using Fingerprint, Face, and Speech.

[8] Samir Akrouf, Belayadi Yahia, Mostefai Messaoud and Youssef chahir, "A Multi-Modal Recognition System Using Face and Speech" in IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 3, No. 1, May 2011

[9] M. Farrús, A. Garde, P. Ejarque, J. Luque, J. Hernando, "On the Fusion of Prosody, Voice Spectrum and Face Features
for Multimodal Person Verification", in Conference: INTERSPEECH 2006 - ICSLP, Ninth International Conference on Spoken Language Processing, Pittsburgh, PA, USA, September 17-21, 2006

[10] Ibiyemi T.S, Aliu S.A, "Face and Speech Recognition Fusion in Personal Identification",in International Journal of Computer Applications (0975 – 8887)Volume 47– No.23, June 2012

[11] J.Deny, Dr.M.Sudhararajan, "Efficient Methods of Multimodal Biometric Security System- Fingerprint Authentication, Speech and Face Recognition",in International Journal of Electrical and Electronics Research ISSN 2348-6988 (online) Vol. 2, Issue 2, pp: (78-83), Month: April - June 2014,

[12] Y. M. Fouda,"Fusion of Face and Voice: An improvement" ,in International Journal of Computer Science and Network Security, VOL.12 No.4, April 2012

[13] Chao-Yo-Lin et al, "User Identification Design by Fusion of Face Recognition and Speaker Recognition", in 12th International Conference on Control, Auto mation and Systems, October 2012

[14] Nagesh Kumar M. and M.N. Shanmukha Swamy, "An Efficient Multimodal Biometric Face Recognition Using Speech Signal"
,2010 IEEE

[15] 15. Tertulien Ndjountche,Rolf Unbehauen, Fa-Long Luo, "The Fusion Framework in a Person Identity Verification System Based on Face and Speech Data", in , 2005 IEEE, CCECE/CCGEI, Saskatoon, May 2005

[16] Sien-Ting Cheng, Yi-Hsiang Chao, Shih-Liang Yeh, Chu-Song Chen, Hsin-Min Wang, and Yi-Ping Hung, Akintola A.G "An Efficient Approach to Multimodal Person Identity Verification by Fusing Face and Voice Information", in ,2005 IEEE