Journal of Basic Sciences
www.skpubs.com
JOURNAL OF BASIC SCIENCES

Multimodal Biometric Recognition System Based on Face and Voice: A Hybrid Approach

Mane B.V[a]., K.V. Kale[b] and A.V. Mane[c]

[a] Department of Computer Science, Sant Gadge Baba Amravati University, Amrawati(M.S.) India.

[b] Department of Computer Science, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad (M.S.), India

[c] Department of Digital and Cyber Forensic, Government Institute of Science, Nagpur (M.S.), India

[*Corresponding Author's E-mail: arjunmane7113@gmail.com]

Abstract:

Multimodal Biometric based recognition system has been extensively studied in the last few decades. In particular the multi trait biometric system. In this paper a multi trait based hybrid approach has been proposed. On the basis of the experiments performed to evaluate the proposed system with traditional multi trait biometric system, it is found that the proposed approach gives better performance.

Keywords: Multimodal Biometric, Face Recognition, Speaker Recognition

1 Introduction

Wide adoption of biometrics has been predicted for years, but has not yet happened mostly because a variety of problems associated with accuracy. Those problems include noise in the sensed data, intra-class variations, distinctiveness, non-universality and spoofing [1]. Keeping in view the bottlenecks attached with uni-modal systems, multi-modal systems offer more robustness in terms of recognition accuracy [2]-[6].

According to the nature of information sources, a multi-modal biometric system can be classified into following categories [1][7]:

- **Multi-sensor systems:** Multi-sensor systems acquire the sample of required modality through multiple sensors or capture units.
- **Multi-algorithm systems**: These systems use different feature extraction and matching algorithms on a single modality acquired through a single sensor. Then, the individual results from each matcher are combined to obtain the final decision.
- **Multi-instance systems**: Such systems capture multiple instances or snap shots of a single modality employing a single imaging sensor from different angles and orientations.
- **Multi-sample systems**: These systems use a single imaging sensor to acquire multiple examples of a single trait.
- **Multi-modal systems**: These systems use multiple uncorrelated biometric signatures acquired through different imaging sensors.

A multi-modal biometric system functions in two modes [1], [7]-[9]:

- **Serial mode:** The acquired multiple traits are processed one after another.
- **Parallel mode**: Multiple modalities are processed simultaneously and the obtained results are combined together to obtain a final match score.

The multiple biometric evidences acquired from different sources can be combined together at different levels [1], [7]-[9]:

- **Sensor level**: The raw data from different sensors is integrated to form a new dataset which is used for feature extraction.
- **Feature level**: The feature sets of different modalities are fused together to form a new feature set, provided that the feature sets must be independent and lie within the same measuring scale.
- **Matching score level**: The feature vectors from different modalities are processed independently and matched with templates through different classifiers. The generated outputs are then fused together and the decision module accepts or rejects the claimed identity based on the composite match score.
- **Decision level**: multiple matchers match the feature vectors with the templates and their decisions are fused together to reach the final decision by employing different techniques such as majority voting

The most commonly used approach in multi-biometric systems is score level fusion [10] because as a quantitative similarity measure it contains rich information about the biometric input, and yet it is still easy to process compared to sensor-level or feature-level data [11].

Unobtrusive human authentication is more convenient than explicit interaction and can also increase system security because it can be performed frequently, unlike the current ‘‘once explicitly and for a long time” practice. Existing unobtrusive biometrics (e.g., face, voice, gait) do not perform sufficiently well for high-security applications, however, while reliable biometric authentication (e.g., fingerprint or iris) requires explicit user interaction. Furthermore, the recognition rates of unobtrusive biometric modalities are highly dependent on environmental conditions: speaker recognition is vulnerable to noise, and video-based biometrics depends on lighting.

This paper introduce the new multi trait based approach for human recognition. In section 2 traditional multi trait biometric recognition system has been discussed. The proposed approach has been discussed in section 3 and experimental results are discussed in section 4. Section 5 conclusion and future scope.

2. Traditional multi trait biometric recognition system

Several researcher works on the face and speech based mult-trait authentication system. But because of the dearth of suitable databases, most face speech fusion studies have been conducted on indoor databases. Various approaches have been explored for score-level and decision-level fusion.

Principal Component Analysis (PCA) has been proven to be an effective technique for automated face recognition [12]-[18] and Mel-Frequency Cepstral Coefficients (MFCC) based approach is most widely used for speaker recognition [19]-[25]. Due to this for the experimentation we choose these two approaches.

2.1. PCA based face recognition

Principal component analysis (PCA) also called as Eigenface [6]. The following steps summarize the process:

1. Let a facial image F(x, y) be a two dimensional mxn array of gray scale intensity values. An image converted as a vector of dimension mn, so that a typical image of size 110x90 becomes a vector of dimension 9900. Let the training set of images {F1, F2, F3…

FN}. The average face of the set is defined by

$$\overline{F} = \frac{1}{N}\sum_{i=1}^{N} Fi \qquad (1)$$

2. Calculate the covariance matrix (C) to represent the scatter degree of all feature vectors related to the average vector. The covariance matrix is computed by

$$C = \frac{1}{N}\sum_{i=1}^{N}(Fi - \overline{F})(Fi - \overline{F})^T \qquad (2)$$

3. The Eigenvectors and corresponding eigenvalues are computed by using characteristic function

$$CV = \lambda V \qquad (3)$$

Where V is the set of eigenvectors associated with its eigenvalue λ.

4. Sort the eigenvector according to their corresponding eigenvalues in descending order.
5. Each of the mean centered image project into eigenspace using

$$W_i = V_i^T (F_i - \overline{F}) \qquad (4)$$

6. In the testing phase each test image should be mean centered, now project the test image into the same eigenspace as defined during the training phase eq (4).
7. This projected image is now compared with projected training image in eigenspace. Images are compared with distance measures. The training image that is closest to the test image will be considered as matched and labelled the same.

2.2. Mel-Frequency Cepstral Coefficients (MFCC) based speaker recognition

MFCC proves more efficient for speaker recognition. Following are the steps for calculating MFCC features of a speech signal [23]-[25].

The first step is Frame Blocking, it is the process of segmenting the input speech signal into a small frame with the length within the range of 20 to 40 msec.

In Step 2 Hamming window is used which integrates all the closest frequency lines.

Step 3 convert each frame of samples from time domain into frequency domain using FFT.

Step 4 is Mel Frequency wrapping, each frame signals are passed through Mel-Scale band pass filter to mimic the human ear.

Step 5 convert the log Mel spectrum into time domain using Discrete Cosine Transform (DCT). The result of the conversion is called Mel Frequency Cepstral Coefficient.

These MFCC will pass to the post processing, i.e. vector quantization is a process which takes large sets of feature vectors and create small set of feature vectors that represent the centroids of the distribution. These feature vectors are grouped to form a codebook for each speaker. In the recognition phase, the data from the unknown speaker is compared to the codebook of each speaker and estimate the difference using Euclidean distance. The decision is to be made on the basis of these differences.

2.1. Multi trait biometric system

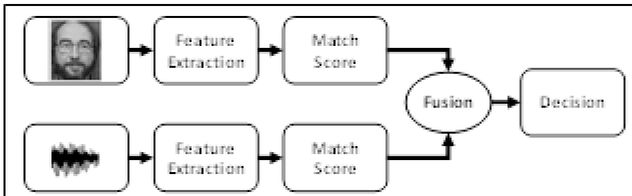

Figure 1. Face and Voice based multi trait biometric

Figure 1 shows block diagram of traditional multi trait biometric system with score level fusion.

3 Proposed method:

Multimodal Biometric Recognition System Based on Face and Voice: A Hybrid Approach has been proposed for human recognition. Figure 2 shows the block diagram of the proposed approach.

The steps of the proposed approach are as follows

1. Let t is the number of training facial images for each class.

2. Create t sets of all training sample, in each set one training facial image of each class.
3. Compute Feature Space of each set using the PCA.
4. Project the test image on each feature space.
5. Calculate the score of testing image on each feature space.
6. For speech signal compute the MFCC feature set and apply the vector quantization to compute the code.

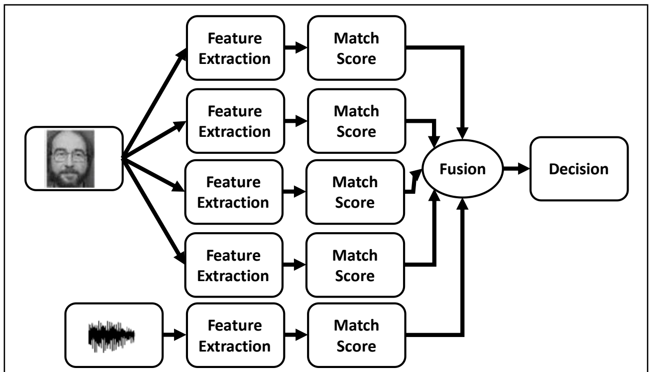

Figure 2. Proposed recognition system

7. Calculate the difference score between training samples and test sample
8. Fuse the score of t projection Step 5 and score from step 7 using min technique.
9. Test image will be labeled with the lowest score class.

4 Experiments

4.1 Facial Database

The ORL (Olivetti Research Lab) facial image database [15] developed by the AT&T Laboratories, Cambridge University has been used for the experiment. It contains slight variations in illumination, facial expression (open/closed eyes, smiling/not smiling) and facial details (glasses/no glasses). Database is of 400 images, corresponding to 40 persons (namely, 10 images for each class). Each image has the size of 112 x 92 pixels with 8-bit gray levels.

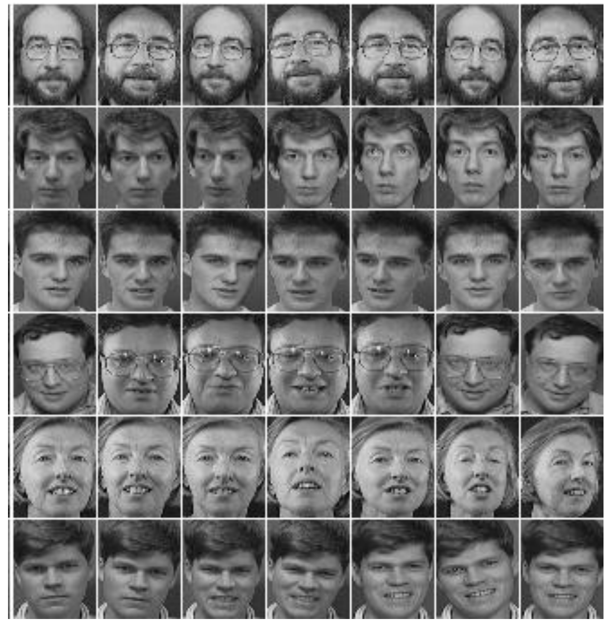

Figure 3. Some images from ORL facial image database

4.2 Speech Database

The KVKRG Voice database developed at Multimodal Biometrics Research Laboratory, Department of Computer Science and Information Technology, Dr. B. A. M. University, Aurangabad

4.3 Experimental Setup

Four experiments have been performed to compare the proposed technique.

In the first experiments the PCA based face recognition technique [6] has been developed and evaluated on the ORL face database of 40 subjects. For training and testing randomly 5 images per subjects has been selected. Second experiment MFCC based speaker recognition technique [23]-[25] has been developed and evaluated on the KVKRG Voice database. For training and testing one sample each has been used. In the third experiments traditional multi-trait system with score level fusion has been developed and evaluated on the mixed database of ORL and KVKRG databases. And Fourth experiment the proposed approach has been evaluated on ORL facial database by reducing size to 40x40, with five number of training images, for testing all the images in the database has been considered and text dependent voice database. All the systems developed in MATLAB R2009a

3 Results and Discussion

The results of the experiments on ORL face database and KVKRG voice database has been shown in TABLE I and the corresponding graphical representation has been shown in Figure 4. From TABLE I, we can analyze that the proposed approach gives the better recognition rate than Unimodal PCA based face recognition system, unimodal speaker recognition system and multi trait biometric system with score level fusion.

Table 1. Experimental Results

	Approach			
	PCA based face recognition	MFCC based speaker recognition	Traditional multi trait approach	Proposed Approach
Recognition Rate	80.50%	60.86%	90%	92.17%

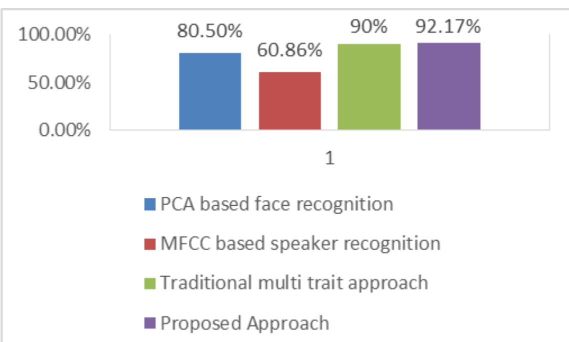

Figure 4. Graphical representation of Table 1.

4. Conclusion and Future Scope

On the basis of experiments we come to conclude that the proposed approach for person recognition gives the improved recognition rate remarkably due to its ability to consider the intra class and inter class variation. In future this work can be extended by evaluating the same approach with different feature normalization and classifier.

5. References

1. A. Ross, K. Nandakumar, A. K. Jain, "Handbook of Multibiometrics", Springer-Verlag edition,2006.
2. Marta Gomez-Barrero, Javier Galbally, Julian Fierrez, "Efficient software attack to multimodal biometric systems and its application to face and iris fusion", Pattern Recognition Letters, Volume 36, 15 January 2014, Pages 243-253.
3. Zengxi Huang, Yiguang Liu, Chunguang Li, Menglong Yang, Liping Chen, A robust face and ear based multimodal biometric system using sparse representation, Pattern Recognition, Volume 46, Issue 8, August 2013, Pages 2156-2168.
4. Heng Fui Liau, Dino Isa, Feature selection for support vector machine-based face-iris multimodal biometric system, Expert Systems with Applications, Volume 38, Issue 9, September 2011, Pages 11105-11111.
5. Le-qing Zhu, San-yuan Zhang, Multimodal biometric identification system based on finger geometry, knuckle print and palm print, Pattern Recognition Letters, Volume 31, Issue 12, 1 September 2010, Pages 1641-1649.
6. Ricardo N. Rodrigues, Lee Luan Ling, Venu Govindaraju, Robustness of multimodal biometric fusion methods against spoof attacks, Journal of Visual Languages & Computing, Volume 20, Issue 3, June 2009, Pages 169-179.
7. J. A. Unar, et al., A review of biometric technology along with trends and prospects, Pattern Recognition (2014), http://dx.doi.org/10.1016/j.patcog.2014.01.016
8. L. P. Rodriguez, A. G. Crespo, M. Lara, B. R. Mezcua, Study of Different Fusion Techniques for Multimodal Biometric Authentication , in : IEEE International Conference on Wireless and Mobile Computing Networking and

Communications, 2008, WIMOB'08 , pp.666–671.
9. M. Gudavalli, A. V. Babu, S. V. Raju, D. S. Kumar. Multimodal biometrics sources, architecture and fusion techniques : an overview. in: Proceedings of the 2012 International Symposium on Biometrics and Security Technologies (ISBAST),IEEE,March2012.pp.27–34.
10. Ross, A.: An Introduction to multibiometrics. In: Proc. Of the 15th European Signal Processing Conference (EUSIPCO), Poznan. Poland (2007).
11. L. Osadciw, P. K. Varshney, and K. Veeramachaneni, "Improving personal identification accuracy using multisensor fusion for building access control applications," in Proc. Fifth Int. Conf. Inf. Fusion, vol. 2, Annapolis, MD, July 2002.
12. M. Turk, "A Random Walk Through Eigenface", IEICE Trans. INF. & SYST., Vol. E84-D, No. 12, Dec. 2001.
13. M. Turk and A. Pentland, "Eigenfaces for recognition", J. of Cognitive Neuroscience, Vol. 3, No. 1, pp. 71-86, 1991.
14. M. Turk and A. Pentland, "Face recognition using eigenfaces", Proceedings of IEEE, CVPR, pp. 586-591, Hawaii, June, 1991.
15. I.Craw, N. Costen, T. Kato, and S. Akamatsu, "How Should We Represent Faces for Automatic Recognition?" IEEE Trans. on PAMI, Vol. 21, No.8, pp. 725-736, August 1999.
16. A. Pentland, B. Moghaddam, T. Starner, "View-based and modular eigenspaces for face recognition", Proceedings of IEEE CVPR, 1994.
17. B. Moghaddam and A. Pentland, "Face Recognition using View-Based and Modular Eigenspace", Automatic Systems for the Identification and Inspection of Humans, 2277, 1994.
18. P. Hall, D. Marshall, and R. Martin, "Merging and Splitting Eigenspace Models", IEEE Trans. on PAMI, Vol. 22, No. 9, pp. 1042-1049, 2002.
19. Campbell, J.P., Jr.; "Speaker recognition: a tutorial" Proceedings of the IEEE Volume 85, Issue 9, Sept. 1997 Page(s):1437 – 1462.
20. 20. Revathi, R. Ganapathy and Y. Venkataramani, "Text Independent Speaker Recognition and Speaker Independent Speech Recognition Using Iterative Clustering Approach", IJCSIT, Vol 1, No 2, November 2009
21. Douglas A. Reynolds and Richard C. Rose, "Robust Text-Independent Speaker Identification using Gaussian Mixture Speaker Models", IEEE Transactions on Speech and Audio transactions and Audio Processing, VOL. 3, NO. 1, JANUARY 1995
22. 22. Alfredo Maesa,Fabio Garzia, "Text independent Automatic Recognition using Mel frequency cepstrum coefficient and Gaussian Mixture Model" , IEEE Proceedings Volume 3,No-4 ,OCT. 2012.
23. F. Bimbot, J. Bonastre, C. Fredouille, G. Gravier, I. Magrin-Chagnolleau, S. Meignier, T. Merlin, J. Ortega-Garcia, D. Petrovska-Delacretaz, and D. Reynolds, "A tutorial on text-independent speaker verification," in EURASIP Journal on Applied Signal Processing, 2004, pp. 430–451.
24. Kavitha K J "An automatic speaker recognition system using MATLAB" World Journal of Science and Technology 2012, 2(10):36-38 ISSN: 2231 – 2587.
25. Kashyap Patel, R.K. Prasad"Speech Recognition and Verification Using MFCC & VQ" International Journal of Emerging Science and Engineering (IJESE) ISSN: 2319–6378, Volume-1, Issue-7, May 2013.
26. The ORL Database of Faces AT&T Laboratories Cambridge http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html