Efficient and Robust Multimodal Biometric System for Feature Level Fusion (Speech and Signature)

Dapinder Kaur
Research Fellow
SGGSWU, Fatehgarh Sahib, Punjab

Gaganpreet Kaur
Research Scholar[1] & Asst. Professor[2]
[1]I.K.Gujral P.T.U, Kapurthala, Punjab
[2]SGGSWU, Fatehgarh Sahib, Punjab

Dheerendra Singh,Ph.D
Prof. & Head
SUSCET, Tangori, Punjab

ABSTRACT

A Pattern can be characterized by more or less rich & varied pieces of information of different features. The fusion of these different sources of information can provide an opportunity to develop more efficient biometric system which is known as Multimodal biometric System. Multimodal biometrics is the combination of two or more modalities such as signature and speech modalities. In this work an offline signature verification system and speech verification system are combined as these modalities are widely accepted and natural to produce. This combination of multimodal enhances security and accuracy. In this work, database can be gathered from 14 users. Each user contributes 4 samples of signature & speech also. Forgeries are also added to test system. 14 forgeries are used for testing purpose. SIFT features are extracted for offline signature which results as a feature vector of 128 numbers & MFCC features are extracted for speech which results as a feature vector of 195 numbers. Fusion at feature extraction level is used in this work by using a new technique named msum which can be proposed by combining sum method & mean method. The experimental results demonstrated that the proposed multimodal biometric system achieves a recognition accuracy of 98.2% and with false rejection rate (FRR) of = 0.9% & false acceptance rate (FAR) of = 0.9%.

General Terms

Multimodal Biometrics, Authentication, msum algorithm for fusion.

Keywords

Biometric, Multimodal Biometrics, Scale invariant features transform (SIFT), Mel Frequency Cepstral Coefficient (MFCC), Feature level Fusion, False Accept Rate (FAR), False Reject Rate (FRR).

1. INTRODUCTION

The need for reliable user authentication techniques has increased in the wake of heightened concerns about security and rapid advancements in networking, communication, and mobility. A wide variety of applications require reliable verification schemes to confirm the identity of an individual requesting their service. Traditional authentication methods using passwords (knowledge-based security) and ID cards (token based security) are commonly used to restrict access to a variety of systems. However these systems are vulnerable to attacked and security can be easily breached. The emergence of biometrics technologies is replacing the traditional methods as it has addressed the problems that plague these systems.

Biometrics refers to the authentication techniques that rely on measurable physiological and individual characteristics that can be automatically verified. Biometric-based solutions are able to provide for confidential transactions and personal data privacy [1]. Multibiometric integrates different biometric systems for verification in making a personal identification. This system takes advantage of the capabilities of each individual biometric. These systems can expect more accuracy due to the fact that they use multiple biometric modalities where each modality presents independent evidence to make a more informed decision. Multimodal biometric systems capture two or more biometric samples and use fusion to combine their analyses to produce a better match decision by simultaneously decreasing the FAR and FRR. All unimodal biometric systems can be used with combination of others to form a multimodal biometrics. For example:

a. Speech and Signature

b. Palm veins & Signature

c. Face & Signature

2. CHOICE OF MODALITY

In this work an offline signature verification system and speaker verification system are combined as these modalities are widely accepted and natural to produce. Although this combination of multimodal enhances security and accuracy, yet the complexity of the system increases due to increased number of features extracted out of the multiple samples and suffers from additional cost in terms of acquisition time [9]. So these days the key issue is at what degree features are to be extracted and how the cost factor can be minimized, as the number of features increases the variability of the intra-personal samples due to greater lag times in between consecutive acquisitions of the sample also increases. Increase in variability of the system will further increase FAR. Thus to resolve these issues an effective feature fusion level is required.

2.1 Level of Fusion

Multibiometric system can be integrated in several different levels as described below [3]:

- Sensor level
- Feature level
- Match score level
- Rank level
- Decision level

Fusion at the match score, rank and decision levels have been extensively studied in the literature. Fusion at the feature level, however, is a relatively understudied. Fusion at this level involves the integration of feature sets corresponding to multiple information sources. Since the feature set contains richer information about the raw biometric data than the match score or the final decision, integration at this level is expected to provide better authentication results. This proposed work presents a novel user authentication system based on a combined acquisition of offline signatures and speech signals. Feature level fusion is used as it is better and gives the optimal identification results.

3. DESIGN & IMPLEMENTATION

This work focuses to implement the Multimodal Biometric System that provides accuracy at limited cost in terms of acquisition time. Each biometric system must perform four basic tasks i.e. acquisition, feature extraction, matching and decision making. Among these the major consideration is on feature extraction. As the number of features increases, the intrapersonal model variability issue arises, which is detrimental to system performance and chances of forgery will also increase [9].

2.1 Proposed Model

The proposed model focuses on following four objectives which are helpful in improving the efficiency of the system and are practically implemented using MATLAB 7.11.0 environment.

a) To Collect the Signature & Speech Samples in data acquisition.
b) To propose a new algorithm/method for feature level fusion.
c) To modify the algorithm for feature extraction in multimodal biometric system.
d) Compare this technique with the current state of art techniques.

In this proposed work, a database of 72 signatures & speech samples consisting of a Training Set & Test Set is used. Training Set consists of 56 genuine signatures & speech samples from 14 known writers. Each Person contributes 4 samples. Test Set consists of 28 signatures & speech samples which consist of 14 genuine signatures & 14 forged signatures.

The result obtained has decreased the FAR as well as FRR & has increased the system performance. The accuracy of the multimodal biometric system has been increased. The total time taken by the Single Sample for feature extraction, fusion & matching is approximately 3-4 seconds.

2.2 Basic Block Design

A multimodal biometric system constitutes of signature and speech acquiring device for generating digital signals and signatures. Simple system architecture is opted as shown in Fig. 1 where both streams of data using feature extraction and modelling tools are modelled independently. The feature vectors are fused using a proposed technique & obtain a new

feature vector which can be stored in database. After storing all data, matcher can be used to match the new data with existing database & gives the results.
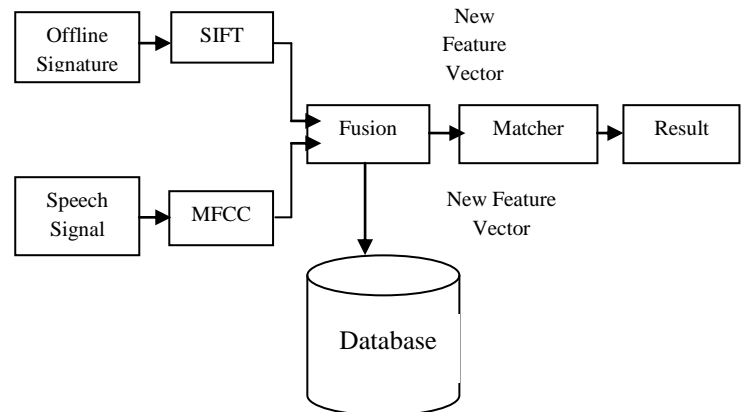

Fig.1 Basic Block Design

2.3 Algorithm Level Design

The algorithm design is shown in Fig. 2, which involves a) Data Acquisition, b) Feature Extraction c) Fusion d) Matching & are discussed in the paper.
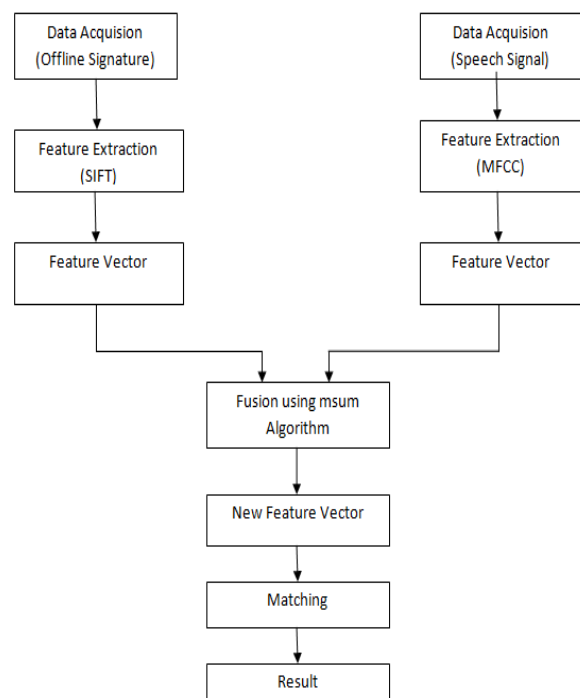

Fig. 2 Proposed Algorithm level Design

2.3.1 Data Acquisition

2.3.1.1 Signature: For offline signature the scanned input is used and for online data acquisition normally the pen tablets have been used. Since off-line handwritten signatures collected via scan images in the different format of image files. So in this work offline signatures are used. For each known person we take a sample of 4 genuine signatures. A database of 72 signatures consisting of a **Training Set & Test Set** is used. Training Set consists of 56 genuine signatures from 14 known writers. Test Set consists of 28 signatures which consist of 14

genuine signatures & 14 forged signatures. Fig. 3 shows some of the offline signature samples gathered by users which are stored in .jpg file format.
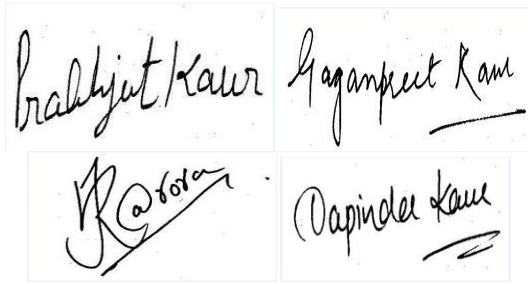

Fig. 3 Offline Signature Samples (.jpg)

2.3.1.2 Speech: To store the different types of speech signals there are multiple equipments but in this work microphone is used to store the voice as .wav file. There are two types of data for speaker recognition text dependent and text independent. In this work, text dependent data has been used for training purpose, and also for testing purpose user has to utter his/her signatures only. A database of 72 speech signals consisting of a **Training Set & Test Set** is used. Training Set consists of 56 genuine speech signals from 14 known persons. Test Set consists of 28 signatures which consist of 14 genuine speech signals & 14 forged signals. Fig. 4 shows some of the speech signal samples gathered by users which are stored in .wav format.
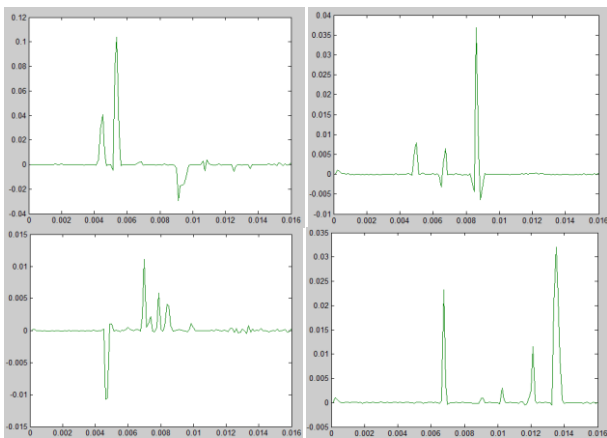

Fig. 4 Speech Signals Samples (.wav)

2.3.2 Feature Extraction

In this work, the features of signature & speech will be extracted by SIFT algorithm and MFCC.

2.3.2.1 Signature: After the data acquisition phase, the signature recognition system then extracts the unique features. Various approaches are possible for signature recognition with a lot of scope of research. In this proposed work, we deal with an Off-line signature recognition technique, where the signature is captured and presented to the user in the format of image only. We use SIFT algorithm to extract the parameters of signature and verify the signature based on these parameters.
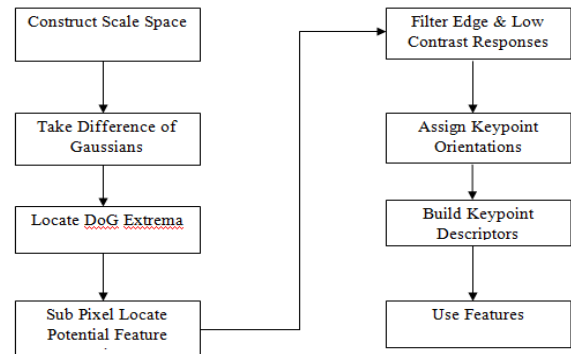
Fig 5 shows the algorithm for computing SIFT features.

Fig. 5 SIFT Algorithm

SIFT algorithm gives 128 no. of feature vector. Signature recognition is a two pattern classification problem, where authentic signatures belong to one class and the forged signatures belong to the other class. Fig. 6 shows the features extracted by using SIFT algorithm on the image file and it shows 128 number of feature vectors in terms of keypoint descriptors.
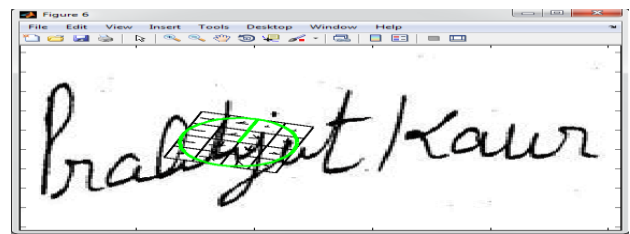

Fig. 6 SIFT Features of Offline Signature

2.3.2.2 Speech: For the speech signal, 12 Mel Frequency Cepstral Coefficients (MFCC) are computed. MFCCs are mean and standard deviation normalized using normalization values computed on the speech part of the data. Fig. 7 shows the algorithm for computing MFCC features of speech signals.
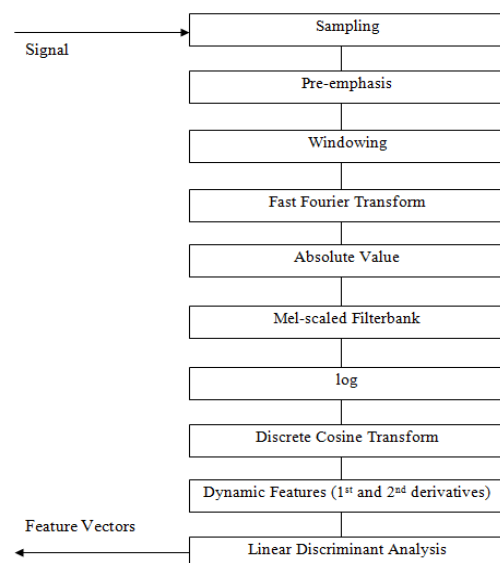

Fig. 7 MFCC Algorithm

MFCC gives 195 no. of feature vector. Fig. 8 shows the feature extracted by using MFCC algorithm on the speech signals and it shows 195 number of feature vectors in terms of cepstral coefficient.
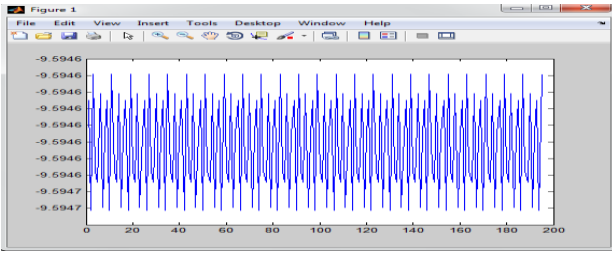

Fig. 8 MFCC Features of Speech Signals

2.3.3 Fusion of Feature Vectors

The main work is fusion at feature level where we have proposed a new algorithm discussed below. In this 128 no. of feature vector can be generated by SIFT algorithm for offline signature & 195 no. of feature vector can be generated by MFCC for speech samples. Both feature vectors are firstly fused with sum method but to add these different dimensional vectors could not be possible. So, padding method can be used where 0 can be added to the end of the lower dimensional feature vector. After fusing using sum method, a new feature vector can be generated which then normalized by mean to compute a single value.

msum Algorithm

a) The feature vectors of signature & speech samples are brought to same dimension. For this:
   - The extra 0 bits are padded to the lower dimensional sample (here signature sample has lower dimensional value than speech signal sample).
b) The sum of feature vectors of signature & speech sample is computed.
   - The sum of two $m \times n$ matrices **A** and **B**, denoted by **A** + **B**, is again an $m \times n$ matrix computed by adding corresponding elements:

$$A + B = \begin{bmatrix} a_{11} & a_{12} & ... & a_{1n} \\ a_{21} & a_{22} & ... & a_{2n} \\ \vdots & \vdots & ... & \vdots \\ a_{m1} & a_{m2} & ... & a_{mn} \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} & ... & b_{1n} \\ b_{21} & b_{22} & ... & b_{2n} \\ \vdots & \vdots & ... & \vdots \\ b_{m1} & b_{m2} & ... & b_{mn} \end{bmatrix}$$
$$= \begin{bmatrix} a_{11} + b_{11} & a_{12} + b_{12} & ... & a_{1n} + b_{1n} \\ a_{21} + b_{21} & a_{22} + b_{22} & ... & a_{2n} + b_{2n} \\ \vdots & \vdots & ... & \vdots \\ a_{m1} + b_{m1} & a_{m2} + b_{m2} & ... & a_{mn} + b_{mn} \end{bmatrix}$$

c) The mean of computed sum of feature vectors is done.
   - x_{ij} is the ith independently drawn observation ($i=1,...,N$) on the jth random variable ($j=1,...,K$). These observations can be arranged into N column vectors, each with K entries, with the $K \times 1$ column vector giving the ith observations of all variables being denoted x_i ($i=1,...,N$).
   - The **sample mean vector** $\bar{x}$ is a column vector whose jth element $\bar{x}_j$ is the average value of the N observations of the jth variable:

$$\bar{x}_j = \frac{1}{N} \sum_{i=1}^{N} x_{ij}, j = 1, ..., K.$$

- Thus, the sample mean vector contains the average of the observations for each variable, and is written:

$$\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i$$

2.3.4 Matching with database

After fusing the features of both signature & speech, the values can be stored in the database. For testing purpose, both modalities extract features & then fuse them by using this proposed technique and after fusing the value can be compared with the stored values in the database and give the results.

4. RESULTS

Performance of the biometric systems is measured by their accuracy in identification, which is calculated using false rejection rate and false acceptance rate. As shown in the Table 1, the FAR & FRR are calculated. Tests are run on the dataset of 14 users. Feature vector are generated both for intruder and genuine user, after those feature vectors are fused using new proposed techniques describe in the table. Results are reported in the form of FAR and FRR which are obtained for a different values of threshold. Accuracy is calculated for new proposed techniques, which can be comparing with the accuracy at score level fusion.

For both Signature & Speech Samples:

Total Number of Samples in the database=56

Number of Sample that falsely accepted=5

$$FAR =$$

$$\frac{\text{Total Number of Samples} - \text{Number of Samples that Falsely accepted}}{\text{Total Number of Samples}}$$

So, $FAR = \frac{56-5}{56} = \frac{51}{56} = 0.91\%$

For both Signature & Speech Samples:

Total Number of Samples in the database=56

Number of Sample that falsely rejected=5

$$FRR =$$

$$\frac{\text{Total Number of Samples} - \text{Number of Samples that Falsely rejected}}{\text{Total Number of Samples}}$$

So, $FRR = \frac{56-5}{56} = \frac{51}{56} = 0.91\%$

Accuracy = 100 − (FAR + FRR) %

Accuracy = 100 − (0.91 + 0.91) %

Accuracy = 98.2 %

Table 1: Comparison of the results of score level fusion & Feature Extraction level Fusion

	Score level Fusion [9]	Feature Extraction level Fusion
FRR	2.5%	0.9%
FAR	2.5%	0.9%
Accuracy	95%	98.2%

Fig. 9 shows the comparison of Score level Fusion & Feature Extraction level Fusion which results that feature extraction level of fusion gives more accurate results than the score level fusion. The accuracy of the system is approximately 98.2% which is better than the other Systems.
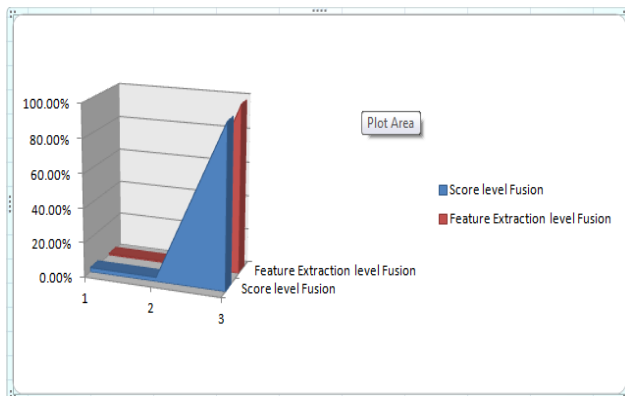

Fig. 9 Comparison of Score level Fusion & Feature Extraction level Fusion

In this work the datasets gathered from 14 persons. Each person contributes 4 samples for both signature & speech. Table 2 shows the time elapsed by system for feature extraction, fusion and matching for 14 genuine samples S1,S2,…..S14 contributed by the users.

Table 2: Elapsed Time

Samples Gathered from 14 users	Time (in sec.)			
	Feature Extraction (Signature+Speech)	Fusion using msum	Matching	Total time
S_1	1.75	0.17	1.54	3.46
S_2	1.72	0.13	1.94	3.79
S_3	1.57	0.13	1.47	3.17
S_4	1.76	0.12	1.83	3.71
S_5	1.82	0.12	1.72	3.66
S_6	1.73	0.12	1.74	3.59
S_7	1.59	0.18	1.83	3.60
S_8	1.65	0.19	1.73	3.48
S_9	1.58	0.12	1.78	3.48
S_{10}	1.75	0.16	1.49	3.40
S_{11}	1.59	0.12	1.74	3.45
S_{12}	1.77	0.12	1.74	3.63
S_{13}	1.84	0.13	1.54	3.51
S_{14}	1.64	0.13	1.77	3.54

5. CONCLUSION & FUTURE SCOPE

In the proposed system a new technique is generated at feature level fusion of offline signature and speech verification system to increase the accuracy of the authentication systems. In this SIFT features are extracted for offline signature & MFCC features are extracted for speech. This proposed method decreased the FAR as well as FRR, & has increases the system performance on the given data set. The accuracy of given system is 98.2%. This system has been compared with the other bimodal system where score level fusion [9] is done using the same modalities. The proposed system has proved that it is more efficient & accurate in terms of accuracy, FAR, FRR and also the time elapsed.

Future works could go in the direction of using more robust modeling techniques against forgeries and hybrid fusion level can be used. Multimodal modalities can be used together to make forgeries more difficult. Also, the system should be tested on a larger database to validate the robustness of the model.

6. REFERENCES

[1] Anil K. Jain, Arun Ross and Salil Prabhakar, "An Introduction to Biometric Recognition," IEEE Transactions on Circuits and Systems for Video Technology, Special Issue on Image- and Video-Based Biometrics,Vol. 14, No. 1, pp.1782-1793, 2004.

[2] Jonas Richiardi and Andrzej Drygajlo, "Gaussian Mixture Models for Online Signature Verification," Speech Processing Group Signal Processing Institute Swiss Federal Institute of Technology (EPFL), WBMA'03, Berkeley, California, USA.ACM 1,pp.771-779, 2003.

[3] Jonas Richiardi and Andrzej Drygajlo, "Gaussian Mixture Models for Online Signature Verification," Speech Processing Group Signal Processing Institute Swiss Federal Institute of Technology (EPFL), WBMA'03, Berkeley, California, USA.ACM 1,pp.771-779, 2003.

[4] Hassan Soliman, "Feature Level Fusion of Palm Veins and Signature Biometrics," International Journal of Video & Image Processing and Network Security IJVIPNS-IJENS Vol: 12 No: 01 28.

[5] S. Perumal Sankar, C. N. Dinakardas, "Multimodal biometric Authentication System Based on High level feature fusion approach," ISSN 1450-216X Vol. 84 No. 1, pp. 55-63,2012.

[6] Harbi AlMahafzah, Mohammad Imran and H.S. Sheshadri, "Multibiometric: Feature level fusion," IJCSI, Vol. 9, Issue 4, No 3, July 2012.

[7] Eren Camlikaya, Alisher Kholmatov and Berrin Yanikoglu, "Multi-Biometric Templates Using Fingerprints and Voice," Biometric Technology for Human Identification V, edited by B. V. K. Vijaya Kumar, Salil Prabhakar, Arun A. Ross Proc. of SPIE Vol. 6944, 69440I, 2008.

[8] R. Gayathri, P. Ramamoorthy, "Feature level fusion of palmprint and iris," IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 4, No 1, July 2012.

[9] Mandeep Kaur, Akshay Girdhar and Manvjeet Kaur, "Multimodal Biometric System using Speech and Signature Modalities," IJCA(0975-8887),Volume 5- No. 12, August 2010.

[10] Pradeep K. Atrey, M. Anwar Hossain, Abdulmotaleb El Saddik, Mohan S. Kankanhalli, "Multimodal fusion for Multimedia analysis: a survey," Multimedia Systems 16:345-379,2010.

[11] R. Brunelli, D. Falavigna, "Person identification using multiple cues," IEEE Transactions on Pattern Analysis and Machine Intelligence 1995.

[12] J. Kittler, R. P.W. Duin, "The combining classifier: to train or not to train," in Proceedings of the International Conference on Pattern Recognition, vol. 16, no. 2, pp. 765–770, 2002.

[13] L. Hong and A. K. Jain, "Integrating faces and fingerprints for personal identification," IEEE Trans. PAMI, vol. 20, no. 12, pp. 1295-1307, 1998.

[14] S. Ben-Yacoub, Y. Abdeljaoued, and E. Mayoraz, "Fusion of face and speech data for person identity verification," IEEE Trans. Neural Networks, vol. 10, no. 5, pp. 1065-1075, 1999.

[15] A. Ross, and A. K. JAIN, "Information fusion in biometric," Pattern Recognition Letters, vol. 24, no. 13, pp. 2115-2125, 2003.

[16] Pradeep K. Atrey, M. Anwar Hossain, "Multimodal fusion for multimedia analysis," Multimedia systems 16:345-379, 2010.

[17] P.D.Garje, Prof. S.S. Agrawal, "Multimodal Identification System," (IOSRJECE) ISSN: 2278-2834 Volume 2, Issue 6, Sep-Oct 2012.

[18] David G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," Accepted for publication in the International Journal of Computer Vision, 2004.

[19] Chadawan Ittichaichareon, Siwat Suksri and Thaweesak Yingthawornsuk, "Speech Recognition using MFCC," International Conference on Computer Graphics, Simulation and Modeling (ICGSM'2012) July 28-29, 2012.

[20] Lindasalwa Muda, Mumtaj Begam and I. Elamvazuthi, "Voice Recognition Algorithms using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques," Journal of Computing, Volume 2, Issue 3, March 2010.

[21] Fang Zheng , Guoliang Zhang and Zhanjiang Song, "Comparison of Different Implementations of MFCC," J. Computer Science & Technology, 16(6): 582-589, Sept. 2001.