© 2015 IJEDR | Volume 3, Issue 3 | ISSN: 2321-9939

# Speech and Iris based Multimodal Biometric System

[1]Sakshi Sahore,[2]Tanvi Sood
[1] M. Tech Student , [2] Assistant Professor
[1]Electronics and Communication Department,
[1] Chandigarh Group of Colleges, Ladran, Mohali, India

---

***Abstract*** **-** Biometrics is the science and technology used for human identification and verification for various applications. Biometric systems can be developed using a single biological attribute is called a unimodal biometric system and the one developed using more than one biological attribute is called multimodal biometric system. While multimodal biometric system suffer limitations in the system performance due to noisy data, interclass variations, non-universality and spoof attacks, multimodal systems are proved to be more accurate and reliable due to multiple sources of information. This paper discusses a multimodal biometric system using speech and iris. The paper presents the fusion of speech and iris modalities at score level. The features of speech are extracted using MFCC while that of iris are extracted using SIFT. GA is used to optimize the features if speech and iris and scores are calculated and fused. Hamming distance is used for matching. The system provides favourable FAR and FRR providing better accuracy.

***Keywords* –** biometrics, unimodal, multimodal, MFCC, SIFT, GA, FAR, FRR

---

## I. INTRODUCTION

The term biometrics is derived from the Greek word Bio and Metric. The term biometrics relates to the measurement (metric) of characteristics of a living (Bio) thing in order to recognize a person. Biometrics uses various physiological or behavioral characteristics. Common physiological biometric measurements include fingerprints, iris, face, hand, retina, etc. While common behavior biometric measurements include signature, speech, rhythm, etc. Single biometric systems have limitations like uniqueness, high spoofing rate, high error rate, non-universality and noise [1]. Multimodal biometric identification system is utilized for solving these limitations. Multimodal biometric is the field of pattern recognition research recognizing the human identity based on physical patterns or behavioral patterns of human. Biometric technique provides the separate characteristics of a person which is always prevalent. Thus the benefit to a biometric is that it does not change or lose. Single biometric system may lead to False Acceptance Rate (FAR) and False Rejection Rate (FRR). Biometric recognition systems are innately probabilistic and their performance needs to be assessed within the context of this fundamental and major characteristic. Biometric recognition involves matching, within a tolerance of approximation of observed biometric traits against biological attributes and behaviors of a person.

The performance of a biometric system is influenced by the reliability of the sensor used and the degrees of freedom offered by the features extracted from the sensed signal. Also, if the biometric feature is detected or measured is noisy (eg a fingerprint with a scar or a broken voice by cold), the final score calculated by the adaptation module may not be reliable. This problem can be solved by different biometric traits. Different biometric characteristics are used by these systems [2]. Biometric systems that use more than a physiological or behavioral trait for identification are called multimodal biometric systems. Multimodal biometric systems should be more reliable due to the presence of multiple evidences [3].

A biometric recognition system works on two modes:

*Enrollment mode:* In the enrollment method of a user's biometric data is acquired using a biometric reader and stored in a database. The stored template is marked by a user identity to facilitate authentication.

*Authentication mode:* In the authentication mode, biometric data of a user is acquired again and the system uses it to verify the claimed identity of the user.

Multimodal biometric systems can work in one of the following scenerios as shown in figure 1.

*Multiple sensors:* The information of the same biometric can be acquired by different sensors [4]. The different samples are then processed by the same algorithm and the results are fused to get the resultant algorithm.

*Multiple instances:* The biometric information is extracted from the multiple instances of the same biometric [5].

*Multiple algorithms*: More than one approach/algorithm is used for feature extraction or classification of the same biometric to improve the system performance [6]

*Multiple biometric*: Evidence from the multiple biometric characteristics is taken [7]

*Multiple samples*: Multiple samples are acquired from the same biometric by a single sensor and processed by the same algorithm to obtain the recognition results [8].

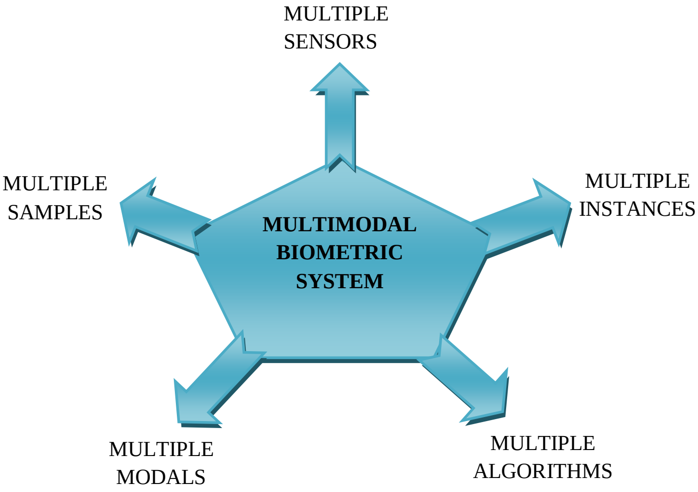


Figure 1: Multimodal Biometric System

Multimodal biometric systems take input single or multiple sensors measuring two or more different ways biometric characteristics. The key to multimodal biometrics is the fusion of different biometric data. A generic multimodal biometric system has four major modules:

1) Sensor level: This fusion strategy requires raw data acquired from multiple sensors that can be subsequently processed and integrated to generate new data from which the features can be extracted. Fusion at the sensor can be that if multiple data of the same biometric is obtained from multiple fit sensors.

2) Feature level: The feature set is extracted from multiple sources of information and is concatenated to form a common feature vector. This new higher dimensional feature vector represents an individual. In feature level matching, reduction technique must be used to select only useful features [9].

3) Score match level: Match Score is a measure of the similarity between the biometric data input and model vectors biometric characteristic. Based on the similarity of feature vector and model, each subsystem calculates its own match the score value. These individual scores are finally combined to obtain a total score, which is then transmitted to the decision module, after which recognition is carried out [10].

4) Decision level: In a multiple biometric system, the merger is done at this level when only the output decisions by individual biometric adapters are available. In this, separate authentication decision is calculated for each biometric feature which is then combined to arrive at a final vote. Various strategies are available to combine the separate decisions of the individual modality to a final authentication decision. Fusion at this level is considered as rigid compared to other merger plans because of the limited availability of information [11].

A number of studies have been done which propose different approaches for multimodal biometrics. Nageshkumar, M., P. K. Mahesh, and MN Shanmukha Swamy presented an authentication method for a multimodal biometric system identification using two traits, face and palm print. The features of palmprint and face features are integrated which increase robustness of the person authentication. The final decision is made by fusion at matching score level architecture. The performance table shows that multimodal system performs better as compared to unimodal biometrics with accuracy of more than 98% [12]. Conti, Vincenzo, Carmelo Militello, Filippo Sorbello, and Salvatore Vitabile proposed an innovative multimodal biometric identification system based on iris and fingerprints performing fingerprint matching using the segmented regions (ROIs) surrounding (pseudo) singularity points while iris preprocessing aims to detect the circular region surrounding the feature, generating an iris ROI. The first test conducted on 10 users resulted in FAR = 0% and FRR = 5.71% while the tests conducted on FVC 2002 DB2A and BATH databases resulted in FAR = 0% and FRR = 7.28% [13]. Nadheen, M. Fathima, and S. Poornima analyzed the performance of dual qualities, ear in addition to iris, independently, and joined them by applying score level fusion procedure. The features were using Principal Component Analysis (PCA) technique by determining the Eigen vectors for 6 dimensionality reduction without information loss. The authors used sum rule based score level fusion method. The proposed system was implemented to study and analyze the performance of multi traits during fusion on a database of sixty users and accuracy of 95% was obtained [14]. Kjhuh A novel multibiometric system using two most used biometric traits fingerprint and iris. The authors extracted the feature vector of each of traits from texture pattern of biometric images, using discrete wavelet transform and principal component analysis technique. The authors classified these feature vectors using Euclidean distance and the he match score obtained were fused together using simple sum rule assigning equal weight to both the modalities. Equal error rate of 0.354500 was achieved for the proposed system [15].

## II. SIMULATION MODEL FOR THE PROPOSED WORK

### *A. Speech recognition*

The goal of this module is to convert the speech waveform, using digital signal processing (DSP) tools, to a set of features (at a considerably lower information rate) for further analysis. This is often referred as the signal-processing front end. The speech signal is a slowly timed varying signal. The general methodology involved in the processing of speech signal includes acquisition of speech data, feature extraction and optimization of features.

The speech data is acquired in .wave format.

*1) Feature extraction*: For feature extraction, different approaches and various kinds of techniques were proposed with varying success rates. The features can be extracted either directly from the time domain signal or from a transformation domain depending upon the choice of the signal analysis approach. Some of the audio features that have been successfully used for audio classification include Mel-frequency cepstral coefficients (MFCC), Linear predictive coding (LPC), Local discriminant bases (LDB).

MFCC is based on the human peripheral auditory system. The human perception of the frequency contents of sounds for speech signals does not follow a linear scale. Thus for each tone with an actual frequency t measured in Hz, a subjective pitch is measured on a scale called the 'Mel Scale' .The mel frequency scale is a linear frequency spacing below 1000 Hz and logarithmic spacing above 1kHz.As a reference point, the pitch of a 1 kHz tone, 40 Db above the perceptual hearing threshold, is defined as 1000 Mels [16].

A block diagram of the structure of an MFCC processor is given in Figure 8. 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-todigital 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.

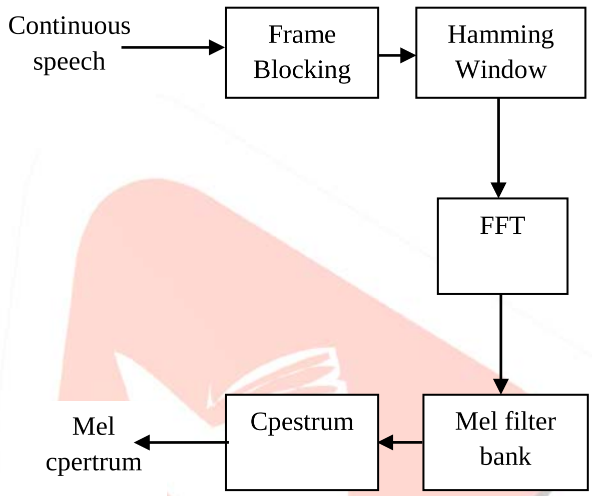


Figure 2: Block diagram for MFCC

*2) Genetic Algorithm:* The optimization of the features refers to the task of identification and selection of a useful subset of features from a larger set of mutually redundant features. The purpose of using GA is to identify a set of features from small or high dimensional data to improve classification accuracy. Genetic Algorithms exert a search in complex, wide landscapes and multimodal and provide near-optimal solutions for fitness or objective function of an optimization problem [17]. In the GA its parameters of the search space are in the coded strings, called chromosomes. A collection of these channels constitute inhabitants. First, a random population is created, the various points in the search space. An objective function and physical fitness is any string that connected the level of quality of the chain is. Based on the principle of survival of the fittest, some strings are selected and assigned to each a certain number of copies. In the opposite pool biologically operators crossover and mutation inspired as Applied to work are chains to give a new generation of chains. The process of selection, crossover and mutation continuously for a fixed number of generations or until the end of the year requirement is met.

### *B. Iris Recognition*

Compared to other biometric features, iris recognition of personal based authentication can get high accuracy due to the rich texture patterns of the iris. Various studies show that the iris is so unique that no two irises are alike, even among identical twins in the general population. Thus, iris recognition has many potential applications such as access control, network security, etc. Iris recognition process consists of various steps, namely image acquisition, image preprocessing, feature extraction and feature optimization.

*1) Image acquisition*: During picture acquisition/capturing of image is the first step of our proposed technique which is collected online. Captured image is of size 10-12 kb and of any format like png, bmp, jpej and so on. The picture acquisition arrangement is shown in below figure.

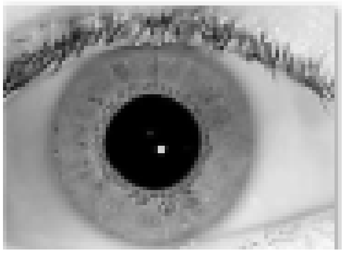

Figure 3: Iris image

*2) Pre-processing:* Image may consist of unusual parts, so it is needed to cut the useful part then forward it to further processing. It is called pre-processing. Preprocessing is divided into three stages. Firstly, the acquired image is converted into a gray scale image. Once the image is gray scaled, the edges of the image are calculated using canny edge detector. Canny edge detector uses a multi-level algorithm to detect the edges. The purpose of edge detection in general is to significantly reduce the amount of information in a picture, although conserving the structural assets to be used for further image processing. After this, HTC is applied. The basic idea of Hough Circular Transform (HCT) is to find curves that can be configured as straight lines, polynomials, circles, etc., in a suitable parameter space. The transformation is able to overcome artifacts such as shadows and noise [18]. The approach is considered good, especially in the face of all kinds of difficulties, including severe occlusions. The outer iris and the pupil are detected through HCT.

*3) Feature extraction using SIFT:* the features of the iris images are extracted using Scale Invariant Feature Transform (SIFT) [19]. SIFT feature extraction consists of the following steps.

*a) Key point detection:* The scale space of an image is defined as a function

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \tag{1}$$

Where * is the convolution operation in x and y, and $G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{\frac{-(x^2+y^2)}{2\sigma^2}}$, $(x, y)$ is the space coordination, σ is the scale coordination, $I(x, y)$ is an input image. To efficiently detect stable key point locations in scale space, we use scale-space extrema in the difference-of-Gaussian (DOG) function convolved with the image $D(x, y, \sigma)$, which can be computed from the difference of two nearby scales separated by a constant multiplicative factor k:

$$D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) \tag{2}$$
$$= L(x, y, k\sigma) - L(x, y, \sigma) \tag{3}$$

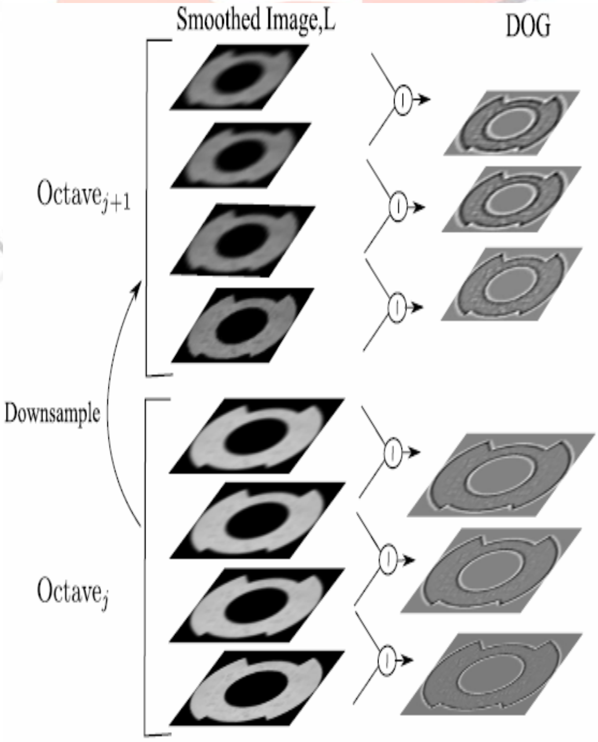


Figure 4: Detection of scale space extrema

*b) Accurate Key-point Localization:* To detect the importance points, DOG pictures are utilized also local maxima as well as local minima are computed across different scales. Each pixel of a DOG image is compared to 8 neighbors in the same scale and 9 neighbors in the neighboring scales.

After key point detection, the next step is performing the detailed fit to the adjoining data intended for location, the proportion of principal curve as well as the scale. The basic idea behind this is to reject all those key points which are low in contrast. These low contrast key points are not considered because as stated in, such key point are sensitive to noise or badly limited to a small area.

*c) Orientation assignment:* To attain invariance to picture rotations, an orientation is allocated towards each and every one of the key-point localities. The descriptor could possibly be represented comparative to this orientation. For determination of the key point orientation, a gradient orientation histogram is worked out in the neighborhood of the key point. A Gaussian smoothed image L1is selected using the scale of a particular key-point. This is followed by formation of the orientation histogram for gradient orientation around each of the particular key-points. The actual histogram encompasses 36 bins designed for 360 orientations and before adding it to the actual histogram, each and every example is weighted by means of gradient magnitude and Gaussian weighted circular frame, by means of σ of 1.5 times the scale of actual key-point. Peaks in histogram correspond to the orientations.

*d) Key-point descriptor:* In this stage, a particular descriptor is registered at every key-point. The picture gradient magnitudes and introductions, with respect to the significant introduction of the key point, are inspected inside a 16X16 locale around every key-point. These specimens are then amassed into orientation histograms summarizing the contents over 4X4 sub regions.

*e) Trimming of false matches:* The key-point matching procedure described may generate some erroneous coordinating focuses. We have evacuated spurious coordinating focuses using geometric limitations. We constrain ordinary geometric varieties to small rotations and displacements. Therefore, if we place two iris images side by side and draw matching lines true matches must appear as parallel lines with similar lengths.

*4) Genetic algorithm:* For the optimization of the iris features, GA is applied.

### *C. Score level fusion*

As discussed earlier, fusion in multimodal biometric systems occur at various levels. Score level fusion is a simple form of fusion which simply integrates the scores obtained from each modality to generate a final score. The similarity score is generated independently each for speech and iris by calculating hamming distance between their feature vectors. Once, independent scores are generated sum rule based fusion is applied and the matching scores of iris and speech are summed up together to form a final fused score.

$$f_{score} = S_{speech} + S_{iris} \quad (4)$$

Where, $f_{score}$ is final fused score, $S_{speech}$ is speech similarity score and $S_{iris}$ is iris similarity score.

## III. RESULTS

This part summarizes the implementation of the proposed system with results. The proposed system is trained for a database of 9 participants. The computational parameters considered are

1) False Acceptance Rate (FAR): It is the measure of the number falsely accepted subjects. It is calculated as
   FAR = total number of subjects – number of falsely accepted subjects / total number of subjects
2) False Rejection Rate (FRR): it is the measure of the number of falsely rejected subjects. It is calculated as
   FRR = total number of subjects – number of falsely rejected subjects / total number of subjects
3) Accuracy: The overall accuracy of the system can be calculated as

$$Accuracy = 100 - (FAR + FRR) \quad (5)$$

Table 1 shows the FAR and FRR of the fused data of speech and iris and figure 5 and figure 6 shows FAR and FRR values against the respective subjects.

Table 1: FAR and FRR of the fused data

| Subjects | FAR | FRR |
|---|---|---|
| 1. | 0.00768 | 0.00456 |
| 2. | 0.00456 | 0.0076 |
| 3. | 0.0093 | 0.0056 |
| **Subjects** | **FAR** | **FRR** |
| 4. | 0.00676 | 0.0084 |
| 5. | 0.00654 | 0.0064 |
| 6. | 0.00789 | 0.0078 |
| 7. | 0.00678 | 0.0079 |
| 8. | 0.00793 | 0.0067 |
| 9. | 0.0568 | 0.0082 |
| 10. | 0.00746 | 0.0079 |


Figure 5: FAR values with respect to the subjects

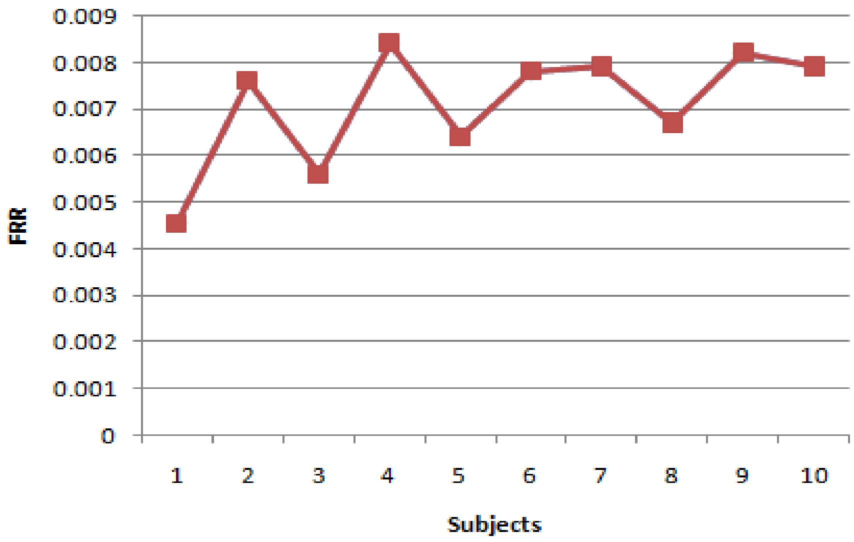


Figure 6: FRR values with respect to the subjects

## IV. CONCLUSION AND FUTURE SCOPE

To overcome the limitations of single biometrics, multimodal biometric technique can be used, which integrates multiple biometric data to make personal identification. In this paper, a novel approach for the fusion of two biometric traits, speech and iris at the matching score level is presented.

The speech signals are processed by using Mel Frequency Cpestral Coding (MFCC) which extracts the features. MFCC converts signals in frequency domain by framing the signals, windowing them and then applying FFT. To reduce the dimensionality, filter banks are used which follows mel-scale. DCT is taken for compressing the information. The iris images on the other hand, are first preprocessed which includes gray scale conversion of the images, edge detection and HTC which detects the pupil boundary and outer iris. Then, the features are extracted using SIFT algorithm. Genetic Algorithm is applied on speech and iris features for selection of a useful subset of features. At last hamming distance is used to calculate the independent similarity scores of speech and iris which are finally fused together to generate the final fused score.

For an ideal authentication system, FAR and FRR should be equal to zero. The proposed fusion of speech and iris at score level is tested on a database of 10 subjects. The accuracy obtained is %. The future work aims to test the system on a larger database and to compare the obtained results with the unimodal systems of speech and iris.

## V. ACKNOWLEDGEMENT

I would like to take this opportunity to express my deepest gratitude to my project supervisor Ms Tanvi Sood (Assistant Professor) for her constant support and guidance throughout this work, my Head of Department Dr. Ruchi Singla and Professor Dr. P.S Mundra for their guidance. Lastly, I would like to thank others who I may have left out for their help and encouragement.

## REFERENCES


[1] P. U. Lahane and S. R. Ganorkar, “Fusion of Iris & Fingerprint Biometric for Security Purpose”, IJSER, vol. 3, pp. 1-5, 2012.
[2] Sahil Prabhakar, Sharath Pankanti, Anil K. Jain, “Biometric recognition: Security and privacy concerns”, Security and Privacy, IEEE, vol. 1, issue 2, pp. 33-42, 2003
[3] Komal Sondhi, Yogesh Bansal, “Concept of Unimodal and Multimodal Biometric systems”, International Journal of Advanced Research in Computer Science and Software Engineering, vol. 4, issue 6, 2014
[4] Thirimachos Bourlai, Nathan Kalka, Arun Ross, Bojan Cukic, Lawrence Hornak, “ Cross Spectral Face Verification in the Short Wave Infrared (SWIR) Band”, Proc. of International Conference on Pattern Recognition, IEEE, pp. 1343-1347, 2010
[5] Dzati Athiar Ramli, Nurue Hayat Che Rani, Khairul Anuar Ishak, “Performances of Weighted Sum Rule Fusion scheme in Multi-instance and Multimodal Biometric system”, World Applied Science Journal, vol. 12, no. 11, pp. 2160-2167, 2011
[6] Vaidehi V., Teena Mary Tressa, Naresh Babu N. T., Annis Fathima A., Balamurali P., Girish Chandra M., “Multi Algorithmic Face Authentication System”, Proceedings of the International Multi-conference of Engineers and Computer Scientists, vol. 1, pp.485-490, 2011
[7] Gandhimathi Amirthalingam, Radhamani G., “A Multimodal Approach for Face and Ear Biometric System”, International Journal of Computer Science Issues (IJCSI), vol.10, issue 5, no. 2, pp. 234-240, 2013
[8] Xi Cheng, Sergey Tulvakov, Ven Govindaraju, “Combination of Multiple Samples Utilizing Identification Modal in Biometric System”, International Conference on Biometric Compendium, IEEE, pp. 1-5, 2011
[9] Vincenzo Conti, Carmelo Militello, Flippo Sorbello, “A Frequency Based Approach for Feature Fusionin Fingerprint and Iris Multimodal Biometric Identification Systems”, IEEE Transactions on Systems, Man and Cybernetics, vol. 40, no. 4, pp. 384-395, 2010
[10] Sayed Hassan Sadeghzadeh, Morteza Amirsheibani, Anseh Danesh Arasteh, “Fingerprint and Speech Fusion: A Multimodal Biometric System”, International Journal of Electronics Communication and Computer Technology (IJECCT), vol. 4, no. 2, pp. 570-576, 2011
[11] Kihal N., Chitroub S., Meunier J., “*Fusion* of Iris and Palmprint for Multimodal Biometric Authentication,” 4th International Conference on Image Processing theory, tools and applications (IPTA), 2014
[12] Nageshkumar, M., P. K. Mahesh, and MN Shanmukha Swamy, "*An efficient secure multimodal biometric fusion using palmprint and face image*", 2009
[13] Conti, Vincenzo, Carmelo Militello, Filippo Sorbello, and Salvatore Vitabile, "A frequency-based approach for features fusion in fingerprint and iris multimodal biometric identification systems", IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, no. 4 384-395, 2010
[14] Nadheen, M. Fathima, and S. Poornima, "Fusion in multimodal biometric using iris and ear", IEEE Conference on Information & Communication Technologies (ICT), pp. 83-87. IEEE, 2013.
[15] S. M. Rajbhoj, and P. B. Mane. "Match score integration of iris and fingerprint in multibiometrics system" IEEE International Conference on Electronics and Communication Systems (ICECS), pp. 1-5. IEEE, 2014.
[16] Chauhan, M. Paresh, and Nikita P. Desai, "Mel Frequency Cepstral Coefficients (MFCC) based speaker identification in noisy environment using wiener filter", International Conference on Green Computing Communication and Electrical Engineering (ICGCCEE), pp. 1-5, 2014.
[17] Wu Jiansheng and Zusong Lu, "A novel hybrid Genetic Algorithm and Simulated Annealing for feature selection and kernel optimization in support vector regression", IEEE Fifth International Conference on Advanced Computational Intelligence (ICACI), pp. 999-1003, 2012.
[18] Nurhanafifi Abd Jalil, R. Sahak and Azilah Saparon. "Iris localization using colour segmentation and circular Hough transform", IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES), pp. 784-788, 2012.
[19] Alonso Fernandez, Fernando, Pedro Tome Gonzalez, Virginia Ruiz Albacete and Javier Ortega-Garcia. "Iris recognition based on sift features”, International Conference on Biometrics, Identity and Security *(BIS)*, pp. 1-8, 2009.