



A New Multimodal Biometric Recognition System Integrating Iris, Face and Voice

Sheetal Chaudhary

Post Doctoral Fellow,

Department of Comp. Sc. & App. K.U.,
Kurukshetra, Haryana, India

Rajender Nath

Professor,

Department of Comp. Sc. & App. K.U.,
Kurukshetra, Haryana, India

Abstract— *The recognition accuracy of unimodal biometric systems has to contend with a variety of problems such as background noise, noisy data, non-universality, spoof attacks, intra-class variations, inter-class similarities or distinctiveness, interoperability issues. This paper describes a new multimodal biometric system that integrates multiple traits of an individual for recognition, which is able to alleviate the problems faced by unimodal biometric system while improving recognition performance. We have developed a multimodal biometric system by combining iris, face and voice at match score level using simple sum rule. The match scores are normalized by min-max normalization. The identity established by this system is much more reliable and precise than the individual biometric systems. Experimental evaluations are performed on a public dataset demonstrating the accuracy of the proposed system. The effectiveness of proposed system regarding FAR (False Accept Rate) and GAR (Genuine Accept Rate) is demonstrated with the help of MUBI (Multimodal Biometrics Integration) software.*

Keywords— *Multimodal Biometric System, Iris recognition, Face recognition, Voice recognition, Score level fusion, Sum rule*

I. INTRODUCTION

A generic biometric system consists of four modules namely sensor module, feature extraction module, matcher module and decision module. In a multimodal biometric system, fusion can be performed depending upon the type of information available in any of these modules. According to Sanderson and Paliwal [1] various levels of fusion can be classified into two broad categories: fusion before matching and fusion after matching as shown in Fig. 1. This classification is based upon the fact that once the matcher of a biometric system is invoked, the amount of information available to the system drastically decreases. Fusion prior to matching includes fusion at the sensor and feature extraction levels and fusion after matching includes fusion at the match score and decision levels. It is generally believed that a fusion scheme applied as early as possible in the recognition system is more effective. The amount of information available to the system gets compressed as one proceeds from the sensor module to the decision module [2].

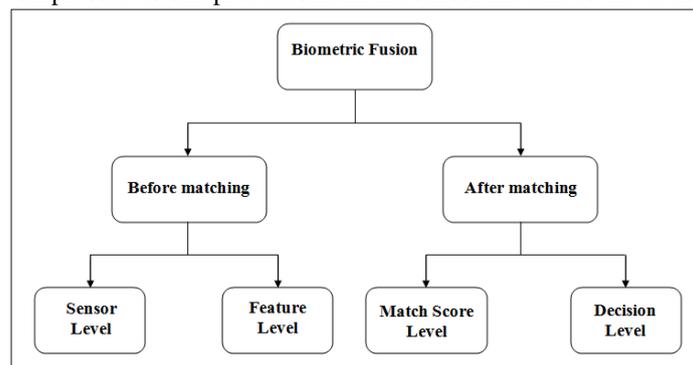


Fig. 1 Classification of levels of fusion

Fusion at the sensor level faces the problem of noise in raw data which gets suppressed in the further levels. Fusion at the feature level involves the consolidation of feature sets corresponding to multiple biometric traits. Since the feature set contains richer information about the raw biometric data than the match score or the final decision, so integration at this level is expected to provide better authentication results. However, it is difficult to achieve integration at the feature level because the relationship between the feature sets of different biometric systems may not be known, the feature representations may not be compatible, concatenating two feature vectors may result in a feature vector with very large dimensionality and a significantly more complex matcher might be required in order to operate on the concatenated feature set [3]. Next to the feature sets, the match scores output by the different matchers contain the richest information about the input pattern and also it is relatively easy to access and combine the scores. Therefore, fusion at the match score level is the most common approach in multimodal biometric systems. Fusion at the decision level contains the least

information i.e. the final output by the system. It is carried out only when the decisions output by the individual biometric matchers are available since most commercial biometric systems provide access to only the final decision output by the system [4].

Integrating multiple traits can significantly improve the recognition performance of a biometric system besides improving population coverage, deterring spoof attacks, and reducing the failure-to-enroll rate. Although the storage requirements, processing time and the computational demands of a multimodal biometric system are much higher than a unimodal system, the above mentioned advantages present a compelling case for deploying multimodal systems in large-scale authentication systems.

The rest of this paper is organized as follows. Section 2 discusses the related work. Section 3 describes the architecture of the proposed system integrating iris, face and voice at the match score level. Results and discussion are given in section 4. Finally, the summary and conclusions are given in last section.

II. RELATED WORK

A lot of work has been done in the last years in the field of multimodal biometrics yielding mature hybrid biometric systems. Fusion at the match score level has been extensively studied in the literature and is the dominant level of fusion in biometric systems.

Luca et al. [5] used fingerprint and face to be fused at the match score level. PCA and LDA are used for the feature extraction and classification. Mean rule, product rule and Bayesian rule are used as the fusion techniques with FAR of 0% and FRR of 0.6% to 1.6%. Kartik et al. [6] combined speech and signature by using sum rule as fusion technique after the min max normalization is applied. Euclidean distance is used as the classification technique with 81.25% accuracy performance rate. Rodriguez et al. [7] used signature with iris by using sum rule and product rule as the fusion techniques. Neural Network is used as the classification technique with EER below than 2.0%. Toh et al. [8] combined hand geometry, fingerprint and voice by using global and local learning decision as fusion approach. The accuracy performance is 85% to 95%. Feng et al. [9] combined face and palmprint at feature level by concatenating the features extracted by using PCA and ICA with the nearest neighbor classifier and support vector machine as the classifier. Fierrez-Aguilar and Ortega-Garcia [10] proposed a multimodal approach including face, a minutiae-based fingerprint and online signature with fusion at the matching score level. The fusion approach obtained Equal Error Rate (EER) of 0.5. Viriri and Tapamo [11] introduced a multimodal approach including iris and signature biometrics at score level fusion with False Reject Rate (FRR) 0.008% on a False Accept Rate (FAR) of 0.01%. Kisku et al. [12] proposed a multibiometric system including face and Palmprint biometrics at feature level fusion. The system attained 98.75% recognition rate with 0% FAR. Meraoumia et al. [13] presented a multimodal biometric system using hand images and by integrating two different biometric traits palmprint and finger-knuckle-print (FKP) with EER = 0.003 %. Aggithaya et al. [14] proposed a personal authentication system that simultaneously exploits 2D and 3D Palmprint features. The sum rule classifier achieves the best EER of 0.002. Kazi and Rody [15] presented a multimodal biometric system using face and signature with score level fusion. The results showed that face and signature based bimodal biometric system can improve the accuracy rate about 10%, higher than single face/signature based biometric system.

III. PROPOSED MULTIMODAL SYSTEM

It is evident that a single biometric trait is not enough to meet the variety of requirements including matching performance and recognition accuracy imposed by several large-scale authentication systems. Multimodal biometric recognition systems appear more reliable due to the presence of multiple, independent pieces of data. They seek to alleviate the shortcomings encountered by unimodal biometric systems by integrating the data presented by multiple biometric traits. In this paper, we develop a fused iris-face-voice recognition system which overcomes a number of inherent difficulties of the individual biometrics. The integrated system also provide anti spoofing measures by making it difficult for an intruder to spoof multiple biometric traits simultaneously.

A. Image Acquisition and Feature Extraction

The images of three traits (iris, face and voice) are acquired using appropriate sensors. The feature extraction of these traits carried out with suitable methods is discussed below:

1) *Iris Feature Set Extraction:* A general iris recognition system is composed of five basic steps: image acquisition, segmentation, normalization, feature extraction and matching. Fig. 2 shows a schematic diagram of these basic steps in the process of iris feature set extraction.

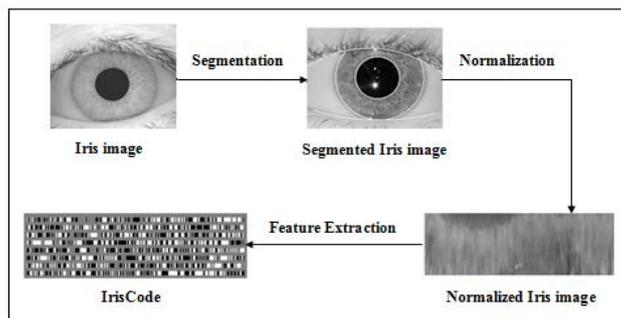


Fig. 2 Steps involved in iris feature set extraction

Segmentation is a process of finding the precise location of the circular iris. The iris region is bounded by two circles. To detect these two circles the Circular Hough transform (CHT) has been used [16]. The size of the iris varies from person to person, and even for the same person, due to variation in illumination, pupil size and distance of the eye from the camera. These factors can severely affect iris matching results. In order to get accurate results, the localized iris is transformed into polar coordinates by remapping each point within the iris region to a pair of polar coordinates (r, θ) where r is in the interval $[0,1]$ with 1 corresponding to the outermost boundary and θ is the angle in the interval $[0,2\pi]$ [17, 18]. Once the iris image has been located, the iris image is encoded into a phase code or IrisCode that is the 2048-bit binary representation of an iris. Gabor filter with isotropic 2D Gaussian function can be used for rotation invariant classification for feature extraction. The matching score is generated by computing the hamming distance between stored IrisCode record with current image. It is a measure of the variation between the IrisCode record for the current iris and the IrisCode records stored in the database.

2) *Face Feature Set Extraction*: Face recognition involves extracting a feature set from a two-dimensional image of the user's face and matching it with the templates stored in the database. The feature extraction process is often preceded by a face detection process during which the location and spatial extent of the face is determined within the given image. To recognize human faces, the prominent characteristics on the face like eyes, nose and mouth are extracted together with their geometry distribution and the shape of the face [19]. Human face is made up of eyes, nose, mouth and chin etc. There are differences in shape, size and structure of these organs, so the faces are differ in thousands ways, and we can describe them with the shape and structure of these organs in order to recognize them. These feature points and relative distances between them make some patterns in every input signal. These characteristic features are called eigenfaces in the facial recognition domain (or principal components). Once the boundary of the face is established and feature points are extracted, the eigenface approach [20] is used to extract features from the face as shown in Fig. 3. In this approach a set of images that span a lower dimensional subspace is computed using the principal component analysis (PCA) technique [21]. The feature vector of a face image is the projection of the original face image on the reduced eigenspace. The matching score is generated by computing the Euclidean distance between the eigenface coefficients of the template and the detected face.

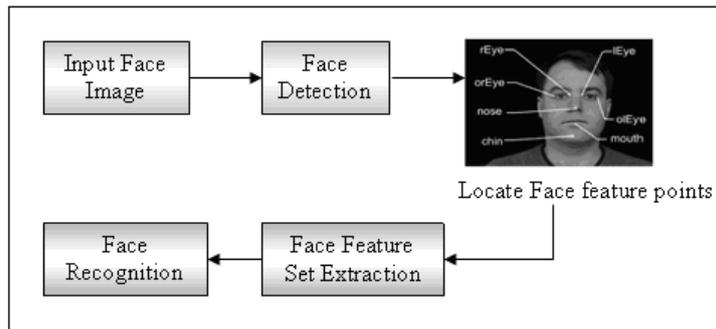


Fig. 3 Steps involved in face feature set extraction

3) *Voice Feature Set Extraction*: Fig. 4 shows a block diagram of steps involved in the process of voice feature set extraction. The Mel Frequency Cepstral Coefficients (MFCC) [22] is one of the most popular feature extraction techniques used in voice recognition based on frequency domain using the Mel scale. MFCC is a representation of the real cepstral of a windowed short-time signal derived from the Fast Fourier Transform (FFT) of that signal [23]. In order to extract the MFCC coefficients the voice sample is taken as the input and hamming window is applied to minimize the discontinuities of a signal. After the windowing, Fast Fourier Transformation (FFT) is calculated for each frame to extract frequency components of a signal in the time-domain. FFT is used to speed up the processing. The logarithmic Mel-Scaled filter bank is applied to the Fourier transformed frame. This scale is approximately linear up to 1 kHz, and logarithmic at greater frequencies. The last step is to calculate Discrete Cosine Transformation (DCT) of the outputs from the filter bank. To enhance the accuracy and efficiency of the extraction process, speech signals are normally pre-processed before features are extracted [24]. The matching score is generated by computing the Euclidean distance between the input signal and template stored in database. The relation between frequency of speech and Mel scale can be established as:

$$\text{Frequency (Mel Scaled)} = [2595 \log(1 + f(\text{Hz})/700)] \quad (1)$$

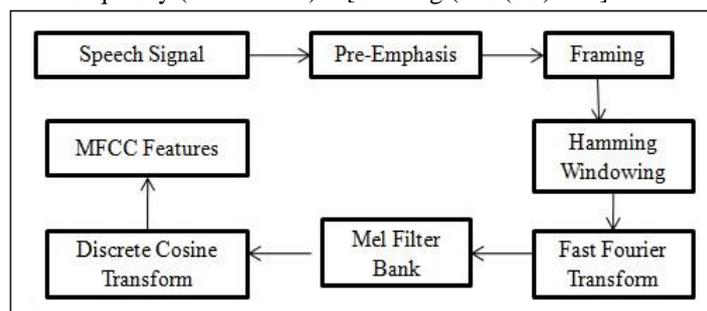


Fig. 4 Steps involved in voice feature set extraction

B. Architecture of Proposed System

The structural design of proposed multimodal biometric recognition system integrating iris, face and voice is shown in Fig. 5. In the operational phase, the three biometric sensors capture the images individually from the person to be identified and converts them to a raw digital format, which is further processed by the feature extraction modules individually to produce a compact representation that is of the same format as the templates stored in the corresponding databases taken during the enrollment phase. The three resulting representations are then fed to the three corresponding matchers. Here, they are matched with templates in the corresponding databases to find the similarity between the two feature sets. The match scores generated from the individual biometrics are then passed to the fusion module to perform fusion at match score level using simple sum rule.

1) *Fusion:* The first step involved in fusion is score normalization. Since the match scores output by the three biometric traits (iris, face and voice) are heterogeneous because they are not on the same numerical range, so score normalization is done to transform these scores into a common domain prior to combining them. Here, min-max normalization is used to transform all these scores into a common range [0, 1]. The three normalized scores are fused using sum rule to generate final match score. Finally, fused matching score is passed to the decision module where a person is declared as genuine or an imposter.

The normalized scores are obtained by following min-max equation [25]:

$$S'_i = \frac{S_i - S_{\min}}{S_{\max} - S_{\min}} \tag{2}$$

where S'_i is the normalized matching score, S_i is the matching score, S_{\min} is the minimum match score and S_{\max} is the maximum match score for i^{th} biometric trait.

In order to combine the match scores output by the three individual matchers (iris, face and voice), simple sum rule is used and its equation is given below [25]:

$$Sum = \sum_{i=1}^n S_i \tag{3}$$

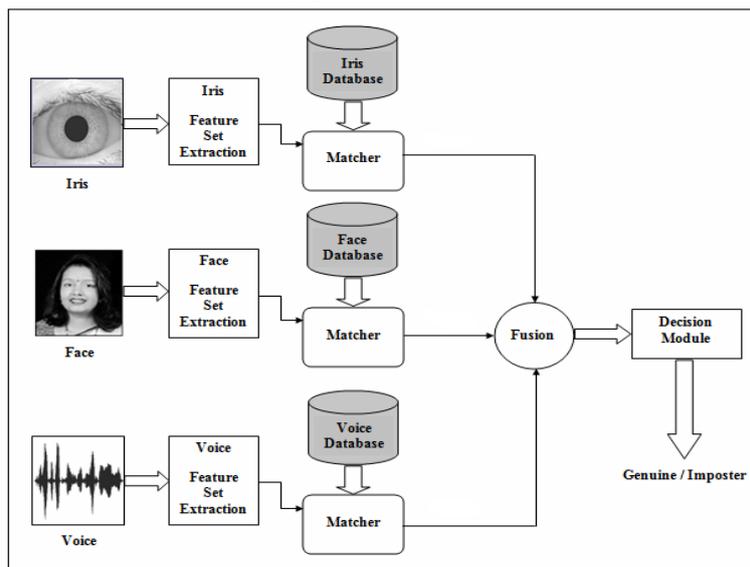


Fig. 5 Architecture of proposed multimodal biometric recognition system integrating iris, face and voice

IV. COMPARISON

This paper presents a new multimodal biometric recognition system integrating iris, face and voice based on score level fusion. The proposed system was implemented using MUBI software. The sample biometric data for iris, face, and voice was taken from CASIA database [26], NIST website [27], and XM2VTS database [28] respectively. Min-Max normalization and simple sum rule fusion strategy were used in the fusion approach of proposed system.

The performance of a biometric system is usually represented by the ROC (Receiver Operating Characteristic) curve. The ROC curve plots the probability of FAR versus probability of FRR for different values of the decision threshold (t). FAR is the percentage of imposter pairs whose matching score is greater than or equal to t and FRR is the percentage of genuine pairs whose matching score is less than t [29]. In order to show the effectiveness of the proposed method, we compare the proposed system with individual biometric traits by plotting Receiver Operating Characteristic curve for Genuine Acceptance Rate (GAR) against False Acceptance Rate (FAR). GAR (1-FRR) is the fraction of genuine scores exceeding the threshold [30]. Fig. 6-8 shows the comparison of proposed system with individual systems on the basis of genuine acceptance rate and false acceptance rate. It can be easily estimated from the ROC curves that the performance gain is very high as compared to the three individual traits. It can also be concluded from Table 1 that the proposed system has improved false acceptance rate as compared to the other individual biometrics. This performance is a significant improvement, even over the best unimodal system (iris) and it underscores the benefit of deploying multimodal systems.

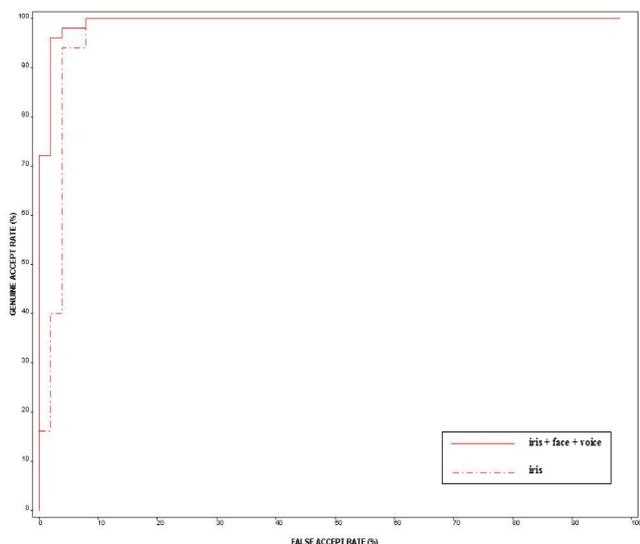


Fig. 6 ROC curve for proposed system and iris

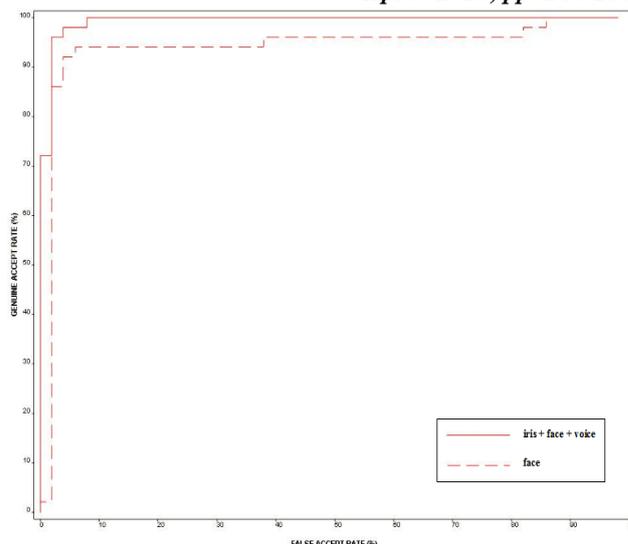


Fig. 7 ROC curve for proposed system and face

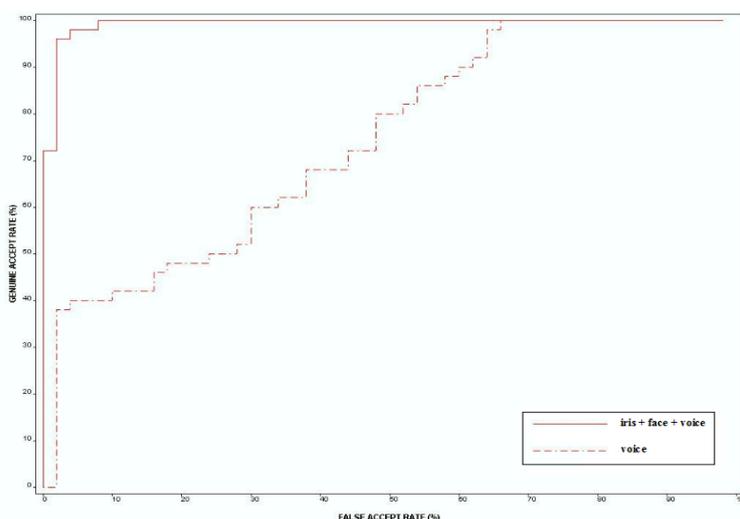


Fig. 8 ROC curve for proposed system and voice

Table I Comparison of proposed system with existing biometrics

S.No.	Biometric Technologies	GAR (%)	FAR (%)
1.	Iris	92	4
	Proposed System		2
2.	Face	92	6
	Proposed System		2
3.	Voice	92	64
	Proposed System		2

V. CONCLUSION

Biometric features are unique to each individual and remain unaltered during a person's lifetime. These features make biometrics a promising solution to the society. In this paper, a robust multimodal biometric recognition system integrating iris, face and voice is proposed. Fusion of three biometric traits is carried out at the match score level. The performance of proposed system is compared with each of the three individual biometrics by plotting ROC curves. These curves show that fusion of multiple biometrics improves the recognition performance as compared to the single biometrics. It also prevents spoofing since it would be difficult for an impostor to spoof multiple biometric traits of a genuine user simultaneously. One of the drawbacks is that database will be very large due to the storage of iris, face and voice template in memory, therefore extra storage space will be needed. Enlarging user population coverage and reducing enrollment failure are additional reasons for combining these multiple traits for recognition. Future work will be focused on integrating liveness detection with multimodal biometric systems and minimizing complexity of the system as it will provide a better solution for increased security requirements.

REFERENCES

- [1] C. Sanderson and K. K. Paliwal, Information Fusion and Person Verification Using Speech and Face, Information. Research Paper IDIAP-RR 02-33, IDIAP, September 2002.
- [2] A. Ross, K. Nandakumar, and A. K. Jain, Handbook of Multibiometrics, New York: Springer, 2006.
- [3] A. Ross and R. Govindarajan, Feature Level Fusion Using Hand and Face Biometrics, In Proceedings of SPIE Conference on Biometric Technology for Human Identification II, volume 5779, pages 196–204, Orlando, USA, March 2005.
- [4] A.K. Jain, A. Ross, Multibiometric systems, Communications of the ACM, Special Issue on Multimodal Interfaces, Vol. 47, January 2004, 34-40.
- [5] Gian Luca Marcialis and Fabio Roli, “Serial Fusion of Fingerprint and Face Matchers”, M. Haindl, MCS 2007, LNCS volume 4472, pp. 151-160, © Springer-Verlag Berlin Heidelberg 2007.
- [6] Kartik.P, S.R. Mahadeva Prasanna and Vara.R.P, “Multimodal biometric person authentication system using speech and signature features,” in TENCON 2008 - 2008 IEEE Region 10 Conference, pp. 1-6, Ed, 2008.
- [7] Rodriguez.L.P, Crespo.A.G, Lara.M and Mezcuca.M.R, “Study of Different Fusion Techniques for Multimodal Biometric Authentication,” in Networking and Communications. IEEE International Conference on Wireless and Mobile Computing, 2008.
- [8] Toh.K.A, J. Xudong and Y. Wei-Yun, “Exploiting global and local decisions for multimodal biometrics verification,” Signal Processing, IEEE Transactions on Signal Processing, vol. 52, pp. 3059-3072, 2004.
- [9] G. Feng, K. Dong, D. Hu and D. Zhang, “When Faces Are Combined with Palmprints: A Novel Biometric Fusion Strategy,” in Biometric Authentication. vol. 307, 2004.
- [10] J. Fierrez-Aguilar, J. Ortega-Garcia, D. Garcia-Romero, and J. Gonzalez Rodriguez, “ A comparative evaluation of fusion strategies for multimodal biometric verification,” in Proc. 4th Int, Conf,Audio-video-based Biometric Person Authentication , J. Kittler and M. Nixon, Eds., vol. LNCS 2688, pp. 830–837, 2003.
- [11] S. Viriri and R. Tapamo, “Integrating Iris and Signature Traits for Personal Authentication using User-Specific Weighting”, 2009.
- [12] D. Kisku, P. Gupta and J. Sing, “Multibiometrics Feature Level Fusion by Graph Clustering”, International Journal of Security and Its Applications Vol. 5 No. 2, April, 2011.
- [13] A. Meraoumia, S. Chitroub and A. Bouridane, “Fusion of Finger-Knuckle-Print and Palmprint for an Efficient Multi-biometric System of Person Recognition”, IEEE ICC 2011.
- [14] V. Aggithaya, D. Zhang and N. Luo “A Multimodal biometric authentication system based on 2D and 3D palmprint features”, Proc. of SPIE Vol. 6944 69440C-1- 2012.
- [15] M. Kazi and Y. Rode, “multimodal biometric system using face and signature: a score level fusion approach” ,Advances in Computational Research, Vol. 4, No. 1, 2012.
- [16] R. Wildes, J. Asmuth, G. Green, S. Hsu, and S. McBride. “A System for Automated Iris Recognition”, Proceedings IEEE Workshop on Applications of Computer Vision, Sarasota, FL, USA, 1994.
- [17] K. Dmitry, “Iris Recognition: Unwrapping the Iris”, The Connexions Project and Licensed Under the Creative Commons Attribution License, Version 1.3. (2004).
- [18] R. Schalkoff., “Pattern Recognition: Statistical, Structural and Neural Approaches”, John Wiley and Sons Inc., pp. 55-63 (2003).
- [19] Dirk Colbry, George Stockman, and Anil Jain, “Detection of Anchor Points for 3D Face Verification”.
- [20] M. Turk and A. Pentland, “Eigenfaces for recognition,” Journal of Cognitive Neuroscience, vol. 3, no. 1, pp. 71–86, Mar. 1991.
- [21] Lu, X.; Wang, Y. & Jain, A.K. (2003), “Combining Classifiers for Face Recognition”, In IEEE Conference on Multimedia & Expo, Vol. 3, pp. 13-16.
- [22] DOUGLAS O'SHAUGHNESSY, “Interacting With Computers by Voice: Automatic Speech Recognition and Synthesis”, Proceedings of the IEEE, VOL. 91, NO. 9, September 2003, 0018-9219/03 2003 IEEE.
- [23] Lahouti, F., Fazel, A.R., Safavi-Naeini, A.H., Khandani, A.K, “Single and Double Frame Coding of Speech LPC Parameters Using a Lattice-Based Quantization Scheme,” IEEE Transaction on Audio, Speech and Language Processing, Vol. 14, Issue 5, pp. 1624-1632, Sept-2006.
- [24] TejalChauhan, HemantSoni, SameenaZafar, “A Review of Automatic Speaker Recognition System”, International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-3, Issue-4 September 2013.
- [25] A. K. Jain, K. Nandakumar, & A. Ross, “Score Normalization in multimodal biometric systems”, The Journal of Pattern Recognition Society, 38(12), 2005, 2270-2285.
- [26] Chinese Academy of Sciences, Center of Biometrics and Security Research, Database of Eye Images. <http://www.cbsr.ia.ac.cn/IrisDatabase.htm>
- [27] National Institute of Standards and Technology (NIST), [U.S. Department of Commerce](http://www.nist.gov).
- [28] K. Messer, J. Matas, J. Kittler, J. Luttin, and G. Maitre, “XM2VTSDB: The extended M2VTS database,” in Proc. 2nd Int. Conf. Audio-Video Based Biometric Person Authentication, Washington, D.C., Mar. 22–23, 1999, pp. 72–77.
- [29] A.K. Jain, A. Ross, S. Prabhakar, “An introduction to biometric recognition”, IEEE Transactions on Circuits and Systems for Video Technology, Vol.14, 4-20, 2004.
- [30] Kumar, A., Wong, D.C.M., Shen, H.C., Jain, A.K., “Personal Authentication Using Hand Images”, Pattern Recognition Letters 27(13), 1478–1486 (2006).