Turkish Online Journal of Qualitative Inquiry (TOJQI)
Volume 12, Issue 7 July 2021: 13174 – 13182

A cascaded deep network architecture for multi modal biometric recognition

B.Abraham David[a], K.Shyamala[b], Y.David Solomon Raju[C]

[a]PG Scholar, Department of ECE, HITS, Telangana, India
[b]Assistant Professor, Department of ECE, HITS, Telangana, India
[c]HOD & Associate professor, Department of ECE, HITS, Telangana, India

Abstract

: It is very often to see that the results obtained from a uni-modal biometric system have negative results when the input subject is deteriorated or damaged. To provide more accurate and high security with biometrics, researchers have integrated the unique feature of more than one biometric system resulting in multi modal system. Face and Iris recognition systems yield higher precision and observed to be robust against tampering because of the unique features that are extracted for classification. In recent times, deep learning has gained a lot of importance in computer vision application especially for image classification with high accuracy and similarity index. Multiple layers of convolution operation are performed on the spatial elements of image, extracting the finer details pertaining to it. This makes the deep network more suitable for solving classification problem. This paper presents a cascaded customized deep network architecture that aims to merge the classification scores of multiple biometrics to attain more robust recognition system that is invariant to illumination. The proposed approach is tested and evaluated with two standard datasets and metrical analysis is compared against state of art methods

Index Terms: Deep Networks, Multi- Modal biometric system, Classification, Score based fusion

1. Introduction

An individual can be recognized based on the unique attributes that were extracted from any physical property of the person, these attributes are termed as patterns that are the key elements of any biometric recognition system [1]. These biometric systems provide reliable and sustainable solution for identifying an individual, due to this these systems are adopted in many of access controls in many of high security private and government organizations. The anatomical characters include face, finger, Palm, Iris and skin which are treated as physiological traits that were employed in most of the biometric systems.

In general, any biometric recognition system refers to single information which is termed as Unimodal systems which are cost effective, simpler, however some of these unimodal systems are non-reliable and sometimes the output of these systems are wrong when the inputs are deteriorated. In order, to overcome these shortcomings of the traditional unimodal systems novel multi-modal systems are introduced which are dependent on multiple input data [2]. These multi-modal systems have multiple merits like (i) Low error rate (ii) Availability (iii) Higher degree of freedom (iv) robust against attacks [3].

To merge the decision from multiple modalities a score base fusion approach has to be employed for attaining final classification result. This fusion process plays a vital role in deciding the overall performance of the multi-modality system. This integration process can be performed at different levels like at sensor level [4], feature-level [5][6], score level [7] and decision level [8]. The fusion at sensor level is not preferable as it may cause more redundancy, feature level fusion is sufficient to identify an individual more accurately however its practical implementation is too tricky since the sources from multiple inputs may be non-compatible [9]. In the recent times, the rank based fusion process has attained high interest due to easiness of integration that concentrates in identifying the individual. Due to the limited source of information, the decision based fusion process it too rigid. But the score level fusion process finds more attractive due to the simplicity and the outcomes of the fusion process is accurate. This process doesn't include complex redundant data with optimum information needed for attaining high recognition. This paper presents, an effective way of integrating score based fusion process for multi- modal biometric system. The scores were attained from two different customized deep networks that were designed for classifying different biometric modalities. The paper is organized as follows, section I presents the introduction, need and importance of multi-modal systems. Section 2, presents the related work pertaining the present context of implementation. Section 3 presents, proposed approach of score based fusion process with experimental results that were presented in section 4 and conclusion remarks in section 5.

2. Related Work

This section presents a brief literature on related works that were conducted by several researchers earlier and that related to preset context of work. In [10] authors have discussed about a combinational fusion approach that are relied on dynamic features and scores that are obtained from the classifiers of face and Iris modal systems. These features are combined together in an algorithmic fashion where an optimal threshold is used as a thumb rule for integration. It was presented that this approach attains higher accuracy and likely to be robust against noise attack. In other work presented by Huo and others [11] where a Gabor filter bank features were employed to fuse the multi modal systems. In this work, the features were extracted using Gabor filters with varying angle and scale factors for Face and Iris Images and later histogram statistics were used to convert them into orientation features that are low dimensional and possess the attribute of high differentiability. PCA (Principle Component Analysis) and SVM (Support Vector Machines) were employed to reduce the dimensionality and classify the features at faster rates.

Texture based multi modal system was proposed by Basma et.al [12], in their work Log Gabor filters were employed to extract the features and combines Spectral regression with kernel based discriminant system. These features are extracted for Face and Iris modal systems and attained an average classification accuracy of 90.75 percent. A new deep learning based approach was proposed by Veeru et.al [13], in this work the authors have included Forward error codes to minimize the false positives and converted a Pseudo feature space of features that were extracted with CNN (Convolutional Neural Network) are integrated using joint representation layer. Finally, the sorted features are selected who dimensionality is reduced, this later creates a binary template which is obtained with distance transform for error correcting codes. In this work the authors have employed Reed Solomon Codes for FEC and pass the generated template through an appropriate decoder which is closest code word which is later forwarded to final template.

Barni et.al [14], have presented a secure and effective multi modal system that aimed to increase the efficiency of the multi modal system by operating them with encrypted codes. Xu et.al [15] have proposed a 3-modality based biometric system with deep learning mechanism. In this work the authors have developed a customized CNN structure for each uni-modal system and later they have integrated them in algorithmic fashion. The work presented in this paper is adapted from the works of Xu. With this multi-modality architecture, the authors have attained an average classification accuracy of 98.5 percent however the approach is tricky and consumes lot of time. So in this work, these two factors are concentrated where the present works aims to design and develop a new architecture and integration process which is simple, faster and yields higher classification rate. Few more networks were presented in [16] [17] that aims to fuse Face, speech and Iris modalities.

3. Face-Iris Multi-Modal System

In this work two modalities like Face and Iris are considered for implementing the proposed score based fusion process. These modalities are very often used in many access control and authentications systems (Eg: Adhaar –Indian Unique Identification System). These modalities are chosen because of the reliability, uniqueness and robust against tampering.

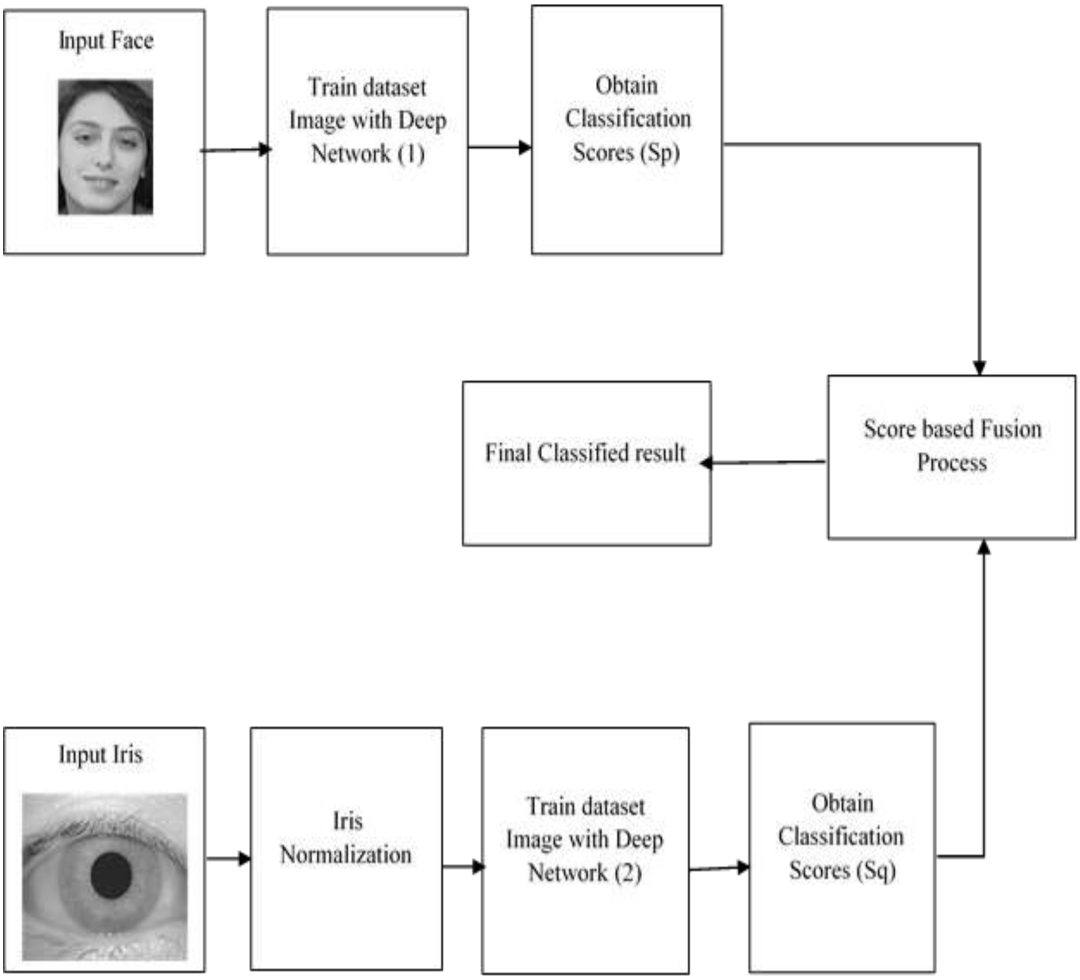


Figure 1: Proposed Score based fusion process block diagram

The proposed approach block diagram is depicted below in figure 1. The proposed approach takes the inputs from two acquisition blocks each for one individual modality. The face images were considered from ORL face data set where each sample is of dimensions 112x92 gray scale images that are translated and have two categories like "with gasses" and "without glasses".

To increase the number of samples, first the images are rotated and scaled with constants such that for each single image we obtain 12 samples. In similar fashion the Iris images were taken from CASIA V1.0 dataset which are gray scale images with 256x256 dimensions. Same as in the case of face images, the Iris images are also rotated and scaled so that we obtain more number of samples for training the Deep network.

Unlike Face images, the Iris images are pre-processed to extract the Iris localized region from the total Eye image. In order to accomplish this activity, we have processed the images with Daughman's rubber sheet model for localizing and extracting the Iris regions [19]. In the present work an integro-differential operator is employed to partition the eye regions.

Thus, obtained segmented portions of Eye images that contain only Iris regions are processed to deep network for training. In parallel to this, face images are processed for other network. So, in total there will be two networks which are custom designed for two modalities The specifications of the networks are given in table 1 and table 2

Table 1: Network specifications for Face Network

Input Image size	112x92x1
2D-Convolutional Layer	5x5x1-920

Max pooling	2x2 with stride factor 2
Fully Connected	7
Softmax Output	Cross Entropy

Table 2: Network specifications for Iris Network

Input Image size	256x256
2D Convolutional Layer	7x7x1-920
Max pooling	2x2 with stride factor 2
Fully Connected	7
Softmax Output	Cross Entropy

The proposed network can be represented pictorially as shown is figure 2. In this network, it consists of 2-convolutional layers each of size 5x5 with 20 filters with stride factor of 1. Max pooling is employed at pooling stage with stride factor of 2. At the end, a fully connected layer of size 7 is employed that aims to classify the subjects with cross entropy classification. Finally, the feature vector is passed to neuron in fully connected layer that produces a k-dimensional vector s = (s1, s2, .., sk) representing class scores towards ‘k’ subjects.

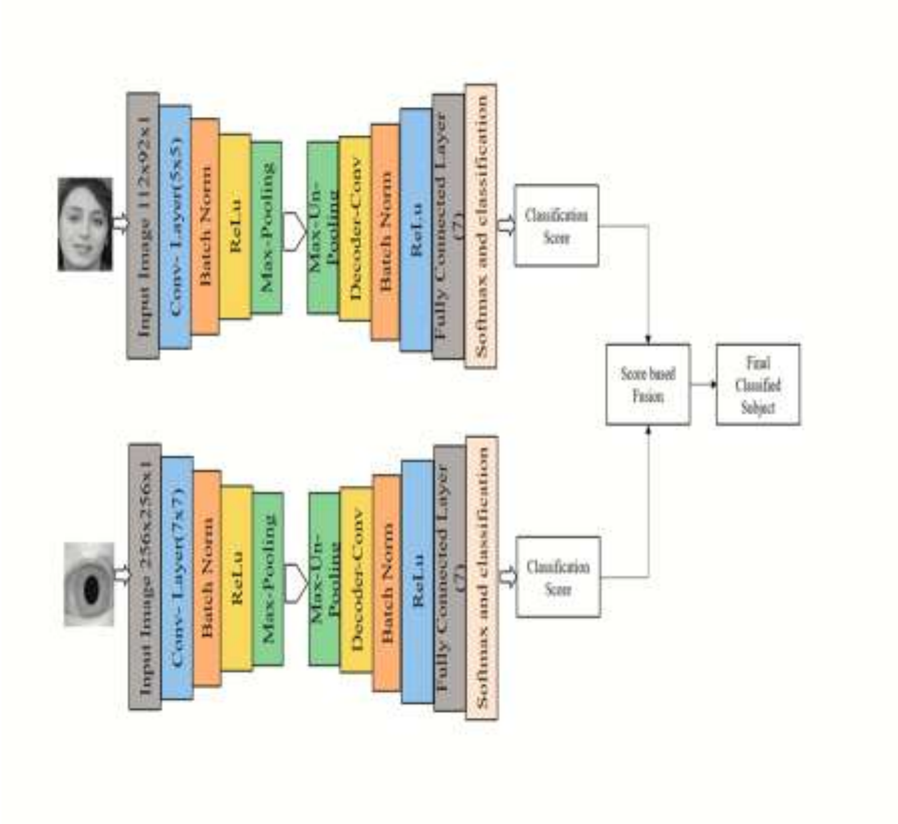


Figure 2: Proposed approach of the network architecture

The proposed score based fusion process is applied to the outcomes of both the networks. It was observed that in many of the cases the scores obtained from individual classifiers are not homogeneous and follow different distribution that may not fall into the same interval [20]-[22]. Let us consider the scores obtained from face network are termed as F(p) and from Iris network as I(q) for the pth and qth subject. The rank of all confidence of all individual scores are calclulated with the following euqation

$$[A(k)]_{(p,q)}=\frac{|s_{(p,q)}(k)-\delta(q)|}{\delta(q)} \qquad (1)$$

In the above equation (1),'k' is number of modalities in this analysis k=2 (Iris, Face), the parameters correspond to pth subject of the qth sample and 'S' corresponds to individual classifier raw score. In brief, if there are 7 subjects each of 12 samples then total number of scores are 84. In the above equation the regularization parameter 'δ' termed as mean-to-overlap extrema-based anchor. The aim of this parameter is to focus on the overlap scores and its neighbours in both the modalities which is given as

$$\delta(q)=\{\max(F(p))-\mathrm{mean}(F(p)\}+\{\mathrm{mean}(I(q))-\min(I(q))\} \qquad (2)$$

Where "F" stands for face score and "I" stands for Iris score.

From the above equation (1) the normalized factor is calculated

$$N(k)=1-\frac{A(k)}{\sum_{k=1}^{m}[A(k)]} \qquad (3)$$

The fusion score (F) is obtained sum product rule as given in equation (4)

$$F=\sum_{k=1}^{m}[N(k)*]\, s_{(p,q)}(k)/m \qquad (4)$$

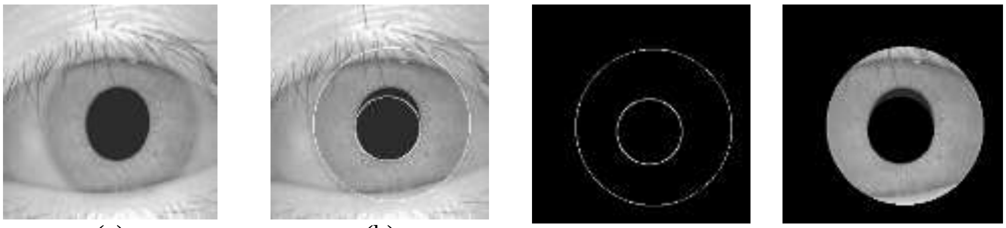
In the above equation 'm' is temporary variable where m=2. Clear numerical analysis in given in the next section.

4. Experimental Results

The experiments of this work were conducted in ORL Face dataset [23], CASIA Iris Dataset (V1.0) [24]. Total 40 out of which 7 subjects were presented with 3 rotation and 3 scaling transformations which are considered for training the network. The Face image is of resolution 112x92 and Iris image dimensions are set to 256x256 gray scale 8-bit.

(a1) (b1) (c1)

Figure 3: a1 Original Image b1 Rotated Image c1: Scaled Images; the images are augmented and used for training the network



(a) (b) (c) (d)

(e)

Figure 4: (a) Original Image (b) Iris Marked using Daughman algorithm (c) Marking or region (d) Extracted Region (e) Normalized Image

For the experimental analysis a confusion matrix was constructed based on the scores of the network.

Table 3: Confusion Matrix of scores (sp,q) for k=1 (Face)

Target (col)/ Predicted score (row)	P1	P2	P3	P4	P5	P6	P7
P1	1	0	0	0	3.49e-13	0	0
P2	4.06e-13	1	0	0	0	3.58e-14	0
P3	0	0	1	2.28e-14	1.22e-21	0	1.46e-12
P4	6.06e-24	0	0	1	4.15e-19	1.69e-13	2.37e-20
P5	1.84e-16	0	7.39e-14	2.9e-24	1	0	2.46e-24
P6	0	2.31e-13	0	3.09e-14	0	1	8.8e-24
P7	0	0	0	0	0	2.10e-14	1

Table 4: Confusion Matrix of scores (sp,q) for k=2 (Iris)

Target (col)/ Predicted score (row)	P1	P2	P3	P4	P5	P6	P7
P1	1	1.41e-16	3.7e-32	2.8e-2	5.3e-19	0	9.8e-35
P2	2.32e-28	1	6.51e-24	0	0	0	9.9e-25
P3	0	0	1	0	1.71e-23	0	0
P4	4.09e-13	0	3.93e-30	1	8.3e-27	0	9.42e-12
P5	1.01e-32	0	9.18e-36	0	1	0	0
P6	6.81e-25	6.91e-41	4.12e-17	48e-34	0	1	9.21e-30
P7	0.0027	1.21e-14	1.48e-8	0.0012	2.46e-5	1.047e-8	0.996

From the above two tables calculate the alpha value. In our experiments alp values are α={1.004, 1,11,1,1,1,.98,1}

Table 5: Confusion Matrix of Fusion scores with Proposed Weighted Normalization

Person/W	P1	P2	P3	P4	P5	P6	P7
P1	0.9997	0	0	0.0129	0	0	0
P2	0	1	0	0	0	0	0
P3	0	0	1	0	0	0	0
P4	0	0	0	1	0	0	0
P5	0	0	0	0	1	0	0
P6	0	0	0	0	0	1	0
P7	0	0	0	0	0	0	1

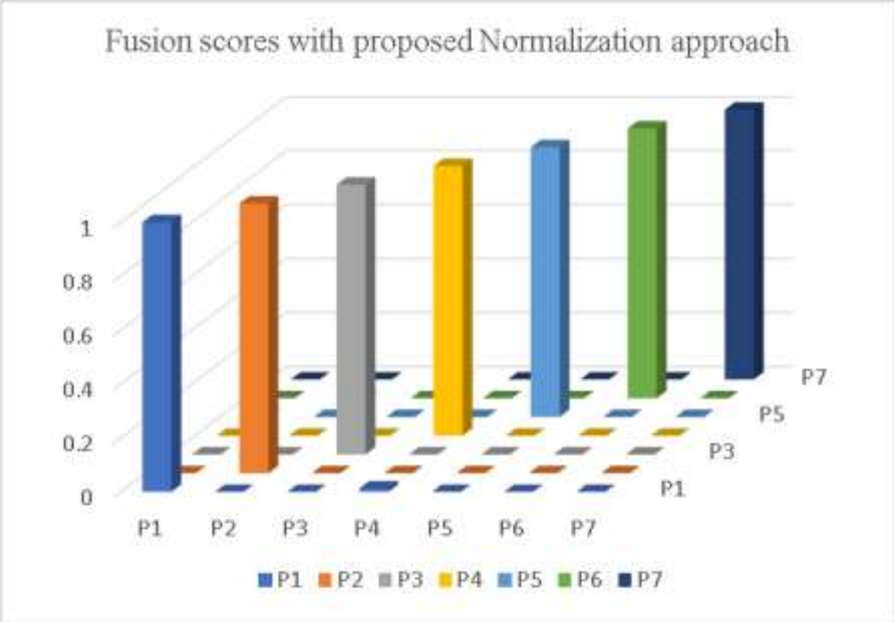


Figure 5: Graphical representation of the proposed weighted Normalization Matrix

The proposed face biometric recognition is compared against face recognition of Huo's method [11] and found that it yields ~98% accuracy which is about 1~1.5% low than the proposed approach.

Table 6: Performance of recognition rates for unimodal and proposed multi modal methods under different geometric transformation

Method/Transformation	**Face –Unimodal**		**Iris-Unimodal**		**Multi Modal**	
	Trained	**Untrained**	**Trained**	**Untrained**	**Trained**	**Untrained**
Normal	1	0.9762	1	0.97	1	0.97
Scaled (70%)	0.897	0.7142	1	0.92	0.945	0.89
Flipped	1	0.8810	0.9821	0.89	0.985	0.9
Average	0.965	0.857	**0.99**	0.92	0.976	**0.92**

Table 7: Performance of correct recognition rates for proposed and Huo's multi modal

Method/Transformation	**Multi Modal [11]**		**Multi Modal (proposed)**	
	Trained	**Untrained**	**Trained**	**Untrained**
CRR	0.983	0.91	1	0.97

Table 8: Performance of correct recognition rates (CRR) for proposed and Huo's multi modal under different geometric attacks

Method/Transformation	**Multi Modal [11]**	**Multi Modal (Proposed)**
Under Illumination scaled to 80%	96.6%	98.5%
Under Gaussian noise of density 10	88.4%	90.2%
Under JPEG compression with quality factor of 30	89.8%	99.05%

Table 9: Performance of CRR of proposed approach and unimodal recognition systems

Method/Transformation	**Multi Modal [11]**	**Multi Modal (Proposed)**
Under Illumination scaled	96.6%	98.5%

to 80%		
Under Gaussian noise of density 10	88.4%l	90.2%
Under JPEG compression with quality factor of 30	89.8%	99.05%

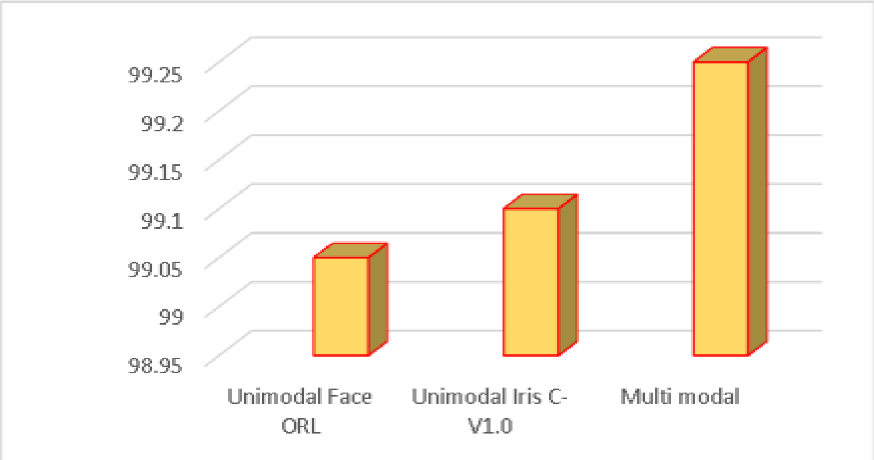


Figure 6: CRR performance comparison for proposed and Unimodal methods

From the results it can be observed that the proposed multi modal weighted normalization approach value ranges between [0-1] that makes all the calculations simpler and which obtains an average efficiency rate of 97.6% under different transformation. In this analysis the data is augmented with rotation and scaling and have not considered the effects like noise, blur and crop. For some Iris image it was observed that the model is mismatching with the imposters, however this is limitation is got through with the proposed weighted normalization.

5.Conclusions

A score based multi modal fusion approach with deep networks is proposed in this paper. This work has considered two bio metrics like face and Iris recognition system for which two individual deep networks were designed and are fused together with sum product rule. In unimodal systems for some inputs a low score was attained leading to mis-match or imposter recognition however this got through with the proposed weighted normalization with which the scores are absolute and attained high accuracy..

References

[1] S. Prabhakar, S. Pankanti,and A. K. Jain” Biometric recognition: security and privacy concerns” IEEE Security and Privacy, pp. 33-42, 2003.

[2] A. Ross, K. Nandakumar, and A. Jain, Handbook of Multi-biometrics. New York, NY, USA: Springer, 2011.

[3] M.R. Alam, M. Bennamoun, R. Togneri, and F. Sohel, “A confidence based late fusion framework for audio-visual biometric identification,” Pattern Recognition Letters, vol. 52, pp. 65–71, Jan. 2015

[4] L. Mezai and F. Hachouf, “Score-level fusion of face and voice using particle swarm optimization and belief functions,” IEEE Transactions on Human-Machine Systems, vol. 45, no. 6, pp. 761–772, Dec. 2015

[5] W.Yang, X. Huang, F. Zhou, and Q. Liao, “Comparative competitive coding for personal identification by using finger vein and finger dorsal texture fusion,” Information Science, vol. 268, pp. 20–32, Jun. 2014.

[6] P. Perakis, T. Theoharis, and I.A. Kakadiaris, “Feature fusion for facial landmark detection,” Pattern Recognition, vol. 47, no. 9, pp. 2783–2793, Sep. 2014

[7] W. Kabir, M.O. Ahmad and M.N.S. Swamy, “A new anchored normalization technique for score-level fusion in multimodal biometric systems,” in IEEE International Symposium on Circuits and Systems (ISCAS), May. 2016, pp. 93–96.

[8] P.P. Paul, M.L. Gavrilova, and R. Alhajj, “Decision fusion for multimodal biometrics using social network analysis,” IEEE Transactions on Systems, Man and Cybernetics: Systems, vol. 44, no. 11, pp. 1522–1533, Nov. 2014.

[9] R. Snelick, U. Uludag, A. Mink, M. Indovina, and A. Jain, “Large scale evaluation of multimodal biometric authentication using state-of the art systems,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, no. 3, pp. 450–455, Mar. 2005.

[10] Maryam Eskandari, Omid Sharifi, “Optimum scheme selection for face–iris Bio-metric”, IET Biom., 2017, Vol. 6 Iss. 5, pp. 334-341

[11] Guang Huo,Yuanning Liu, Xiaodong Zhu, Hongxing Dong,and Fei He, “Face–iris multimodal biometric scheme based on feature level fusion”, Journal of Electronic Imaging, Vol. 24(6), Dec 2015

[12] Basma Ammour, Toufik Bouden, Larbi Boubchir, “Face–iris multi-modal biometric system using multi-resolution Log-Gabor filter with spectral regression kernel discriminant analysis”, IET Biom., 2018, Vol. 7 Iss. 5, pp. 482-489

[13] V. Talreja, M. C. Valenti and N. M. Nasrabadi, "Multi-biometric secure system based on deep learning," 2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP), Montreal, QC, 2017, pp. 298-302

[14] Mauro Barni, Giulia Droandi, Riccardo Lazzeretti, Tommaso Pignata, “SEMBA: secure multi-biometric authentication”, IET Biom., 2019, Vol. 8 Iss. 6, pp. 411-421

[15] H. Xu, M. Qi and Y. Lu, "Multimodal Biometrics Based on Convolutional Neural Network by Two-Layer Fusion," 2019 12th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), Suzhou, China, 2019, pp. 1-6.

[16] K. Prasada Rao, M.V.P. Chandra Sekhara Rao, N. Hemanth Chowdary, An integrated approach to emotion recognition and gender classification, Journal of Visual Communication and Image Representation, Volume 60, 2019, Pages 339-345.

[17] Al-Waisy, A.S., Qahwaji, R., Ipson, S. et al. A multi-biometric iris recognition system based on a deep learning approach. Pattern Anal Applic 21, 783–802 ,2018.

[18] J. Daugman. How iris recognition works. Proceedings of 2002 International Conference on Image Processing, Vol. 1, 2002

[19] Kondareddy, R. Obul and B. Abraham David. “Improved Normalization Approach for Iris Image Classification Using SVM.” (2019).

[20] R. Snelick, U. Uludag, A. Mink, M. Indovina, and A. K. Jain “Large scale evaluation of multimodal biometric authentication using state-of-the-art systems” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, No. 3, pp. 450-455, 2005

[21] R. Cappelli, D. Maio, and D. Maltoni “Combining fingerprint classifier” Proceedings of First International Workshop on Multiple Classifier Systems, pp. 351-361, 2000.

[22] Fouda, Yasser, “Fusion of Face and Voice: An improvement”, IJCSNS International Journal of Computer Science and Network Security, VOL.12 No.4, April 2012

[23] http://cam-orl.co.uk/facedatabase.html

[24] http://biometrics.idealtest.org/findTotalDbByMode.do?mode=Iris