Received: January 4, 2022. Revised: February 15, 2022.

INASS

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Challenge Responsive Multi Modal Biometric Authentication Resilient to Presentation Attacks

Asha Kethaganahalli Hanumanthaiah[1]* **Manjunathswamy Byranahalli Eraiah[1]**

[1]*Department of Computer Science & Engineering, Don Bosco Institute of Technology, Bangalore, India*
*Corresponding author's Email: asha.kh06@gmail.com

Abstract: Multimodal biometric systems are a natural evolution of traditional biometric authentication systems with use of multiple biometric to enhance the security level. Presentation attacks in form of photos and replay videos are becoming a threat to the biometric authentication systems and it becomes necessary to prevent from these attacks for an enhanced security level. This work proposes a challenge response based multi modal biometric authentication resilient to presentation attack. The proposed system is designed for multi modal biometric involving face and iris biometrics with challenge in form of arbitrary emotion invoking image at a random position on the screen. The changes in region of interest around the Facial landmarks and distance of iris to Facial landmarks are captured as response. The response is then matched with expected response derived based on the position of arbitrary emotion invoking image on screen to detect presentation attacks. In addition the work also proposes a hybrid deep feature which provides an addition security in recognition process to ensure authentication failure for fake samples. The error rate in leakage of spoofs is very low in proposed solution (less than 1%) in different environments compared to existing works. The recognition accuracy of face is atleast 1% higher and atleast 2% higher for IRIS in proposed solution compared to existing works.

Keywords: Authentication, Multimodal biometric, Presentation.

1. Introduction

Nowadays traditional authentication mechanisms like token based (cards, keys etc) and knowledge based systems (password, patterns etc) are being replaced with biometric authentication systems. These systems provide various advantages in forms of theft resistance, convenience, uniqueness and higher recognition accuracy. Biometric features have become a stronger and reliable authentication systems compared to traditional counterparts of token based and knowledge based systems. However, the status quo of biometric authentication systems is being recently challenged with presentation attacks in form fake biometric samples with close resemblance to real samples and replay attacks. The fake samples can be created for face; iris and finger prints or replace attack can be launched with printed photos or video playback in front of acquisition devices with sole purpose of deceiving the biometric recognition systems and gaining access to the protected systems. This work addresses the problem of presentation attacks in face and iris based multi modal biometric recognition systems. Both face and iris based authentication are used in various applications like smartphone, computer login, passport control and premises access control. In spite various challenges in acquisition like pose and illumination variations, these two are the most popular biometric authentication systems. Presentation or spoofing attack can be launched on these systems in form of artificial silicone masks, video displays, printed photographs etc.

Various solutions have been proposed in literature for detecting presentation attacks. These solutions user various cues like color, texture, depth, light reflection analysis to detect presentation attacks. The typical biometric authentication system with presentation attack detection works in two

stages. Right after biometric image acquisition, presentation attack detection process analyzes the biometric image. Once the image is detected real, it is passed to the next stage of recognition process. Features set for both the process are quite different. The features are analyzed either by statistical means or using machine learning algorithms.

This work proposes a challenge response based multimodal biometric authentication involving face and iris biometric. An arbitrary emotion provoking image is placed in a random position of screen and presented as challenge to the user. The response of user is captured and compared to expected response to detect spoofing attacks. By this way proposed solution is secure against both photo attacks and video replay attacks. In addition a variant of local binary pattern is designed and integrated into a deep convolutional network to provide a hybrid deep learning feature. Recognition system designed with SVM classifier on the hybrid deep learning features takes the fake samples to negative cases with higher probability. Following are the novel contributions of this work

1. Challenge response based mechanism to detect both photo and video replay based presentation attack.
2. A variant of local binary pattern for biometric system with integration into the deep learning convolutional neural network to get a novel hybrid deep feature for a enhanced security during recognition.

The rest of the paper is organized as follows, in section 2, related works on presentation attack detection systems for biometric authentication systems are discussed. In section 3, the research gaps are detailed and problem is defined. In section 4, the proposed challenge response mechanism to mitigate presentation attack and the novel hybrid deep feature for recognition are detailed. In section 4, the performance results of proposed solution and comparison with state of works are presented. In section 5, the conclusion is presented.

2. Related work

The survey is done in two categories of Face spoof and Iris spoof detection.

Souza et al [1] integrated LBP with convolutional neural network in the first layer for enhanced learning of local pattern features. Though this approach works well for artificially generated images, it fails in case of illumination variations. A new local textural pattern feature called dynamic local ternary pattern (DLTP) is proposed to detect face liveliness by Parveen et al [2]. Though the method works well for face spoof detection, it can be deceived using variations in illuminations. Akhtar et al [3] proposed a method to detect discriminative patches with higher correlation to spoofing attack. The discriminative patches are found by analyzing the in homogeneity in local intensities. The detected discriminative patches are compared statistically to detect spoofing. But the method can be deceived through replay attack. Zhou et al [4] extracted multi scale color features and local directional number pattern features from the image. These two features in combination resist the illumination variations. Though the method detects artificial faces, it fails for replay attacks. Li et al [5] detected the pixel variations due to pulse around the upper facial regions and analyzed it statistically to detect face spoofing. Pixel variations over a time period is collected and power spectral density features are extracted. These features are then thresholded to detect liveliness. The approach was able to detect print attacks but failed for video replay attacks. Hasan et al [6] used contrast and texture features to detect face spoofing. Contrast and texture feature are extracted using Gaussian filtering at different scales. SVM classifier is then used to process feature and detect spoofed photos. But the solution can be deceived with illumination variations. Cai et al [7] extracted meta patterns from face image using a two stream hierarchical fusion network. This pattern along with color featured from the RGB image is used to face liveliness. The method can be deceived with video replay attack. Zheng et al [8] combined depth and multi scale features to detect face spoofing. Higher level feature representation of depth and multi scale is learnt using a two steam spatial temporal network.

Though the method works well for photo attacks, it can be deceived using video replay attacks. Song et al [9] proposed face spoof detection method based on depth cue. Face image acquired from binocular camera is used. From the two images, depth information is extracted and classified using SVM classifier to detect spoofs. But the approach can be deceived presenting two different images in quick succession matching the acquisition order. Cai et al [10] extracted discriminating local features from image, which can be used for spoofing detection. Deep learning convolutional neural network is combined with recurrent neural network is used to detect the discriminating features. The method works best for faces created using tools like Deepfakes but fails for video replay attacks. Yu et al [11] used material characteristics detected from

reflection to classify spoofed images. The material from where image is reflected can be skin, glass, paper or silicone. The reflection pattern is classified using a trained bilateral CNN to detect the material. By this way photo, video reflection or silicone surface etc can be detected. But the method can be disrupted by reflecting noise along with image. Tu et al [12] classified the temporal features using deep learning to detect spoofing. Temporal features are the motion cues in eye, mouth and head. Features are extracted using CNN and classified by the LSTM to discriminate movements for liveliness detection. But method fails in presence of video replays. Wang et al [13] acquired face image in four different modalities and extract deep learning features using Resnet model. Softmax classifier is trained to classify the Resnet features to real or fake face. Though the method detects photo spoofs it fails in presence of video replays. Liu et al [14] used sequence of random light intensities and used the reflection properties to detect spoofs. But this method has higher sensitivity to surrounding artificial light sources which increases the error rate. Chou et al [15] proposed a multi modal approach in which challenge is raised to user and voice response is expected from the user. Based on the voice response, liveliness is detected. But with recent voice tools it becomes easy to speak in someone else voice and deceive the system. Boutellaa et al [16] also used voice modality based challenge to detect spoofing, but this too suffers from same problem of voice faking. Wang et al [17] acquired image from two different camera placed at different places and used the depth information in the acquired image to detect spoofs. Through careful manipulation of image, the solution can be deceived. Agarwal et al [18] proposed a local descriptor for detecting iris spoofing. The descriptor is productive and learns correlation for a pixel with its hexagonal neighbors. Hexagon was selected due to its symmetric properties and higher angular resolution. Bhogal et al [19] proposed six different picture quality measures to assess genuine biometric attributes. But it is easy to create masks to deceive these quality measures. Fathy et al [20] extracted entropy esteems from the wavelet channels and LBP transformations. These entropy esteems are then thresholded to detect fake iris samples. Nguyen et al [21] extracted deep learning features for the iris image captured from the near infrared light camera sensor in various intensities. The detection results for each image are then fused using weighted scoring means to detect iris spoofs. Gavisiddappa et al [31] proposed a multi modal biometric authentication system involving face, iris and finger print. Handcrafted features of LBP, HoG and GLCM are extracted from the biometric image. The features are classified using SVM classifier. But this method did not consider impact of replay attack during authentication. Bouchiba et al [32] used gait and ECG biometric signals for authentication. Their integration to face based biometrics is difficult due to complexity in data acquisition. Abed et al [33] extracted palm vein features and used for authentication. Though the complexity of vein patterns makes it difficult to forge palm vein, the palm vein based authentication system cannot be prevented from replay attack.

3. Problem definition

The summary of survey is presented in Table 1. From the survey, it can be seen that most of the methods for preventing biometric authentication works fail in case of video based replay attacks. Typical presentation attack detection mechanisms based on texture, depth, light reflections, colors, sound etc can be deceived with carefully designed attack plan. The failure is due lack of random challenge mechanism with difficult to create response in a short interval of time (milli seconds). The proposed work addresses this problem.

4. Challenge response multi model biometric reorganization

This work proposes a challenge mechanism which is completely random and difficult for attackers to compute response for challenge in short interval of time.

The proposed solution has two stages: liveliness detection and hybrid deep feature based recognition. Each of the stage is detailed below.

A. Liveliness detection

The challenge is an arbitrary emotion provoking image positioned randomly in the screen. The facial image is collected from camera with eyes in open condition. The facial response and iris position related to facial landmarks are collected. This actual response is matched with expected response to detect the liveliness of the biometric input. Only when liveliness is detection, the second stage of hybrid deep feature based recognition is invoked.

When user is ready for authentication process, camera is set on. An, an arbitrary emotion proving image is placed on screen at random position (x_p, y_p) and Facial images are obtained for certain interval of time. From the sequence of face images, face regions obtained the Facial image using Voila

Table 1. Survey summary

Author	Approach	Drawbacks
Souza et al [1]	LBP integrated into convolutional layer of CNN	Fails in case of illumination variations
Parveen et al [2]	New local textural pattern feature to detect face spoof	Deceived using variations in illuminations
Akhtar et al [3]	discriminative patches are compared statistically to detect spoofing	Fails for replay attacks
Zhou et al [4]	multi scale color features to detect artificial faces	Fails for replay attacks
Li et al [5]	Analyzed pixel variations due to pulse around the upper facial regions to detect face spoof	Able to detect print attacks but failed for video replay attacks
Hasan et al [6]	Contrast and texture feature are extracted using Gaussian filtering at different scales to detect spoofing	Solution can be deceived with illumination variations
Cai et al [7]	Colour and meta features are used to detect face liveliness	Fails for video replay attack
Zheng et al [8]	combined depth and multi scale features to detect face spoofing	Fails for video replay attack
Song et al [9]	face spoof detection method based on depth cue	The approach can be deceived presenting two different images in quick succession matching the acquisition order.
Cai et al [10]	discriminating local features for face spoofing detection	Works only for artificial faces
Yu et al [11]	used material characteristics detected from reflection to classify spoofed images	Accuracy disrupted by reflecting noise along with image
Tu et al [12]	Used motion cues in eye, mouth and head to detect spoofing	Fails in presence of video replays
Wang et al [13]	acquired face image in four different modalities and detected liveliness	Fails for video replay attacks
Liu et al [14]	used sequence of random light intensities and their reflection properties	Higher error due to surrounding artificial light sources
Chou et al [15]	Challenge response in terms of audio feedback	Can be deceived easily
Boutellaa et al [16]	Voice modality based challenge response	Can be deceived using voice faking
Wang et al [17]	Analyzed depth information from image captured using two cameras	Can be deceived with careful manipulation of image
Agarwal et al [18]	local descriptor for detecting iris spoofing.	Fails for video replay attacks
Bhogal et al [19]	six different picture quality measures to assess genuine biometric attributes	easy to create masks to deceive these quality measures
Fathy et al [20]	Thresholded entropy esteems to detect iris spoofing	Fails in case of illuminations
Nguyen et al [21]	Multi modal acquiring iris acquisition with varying light intensities.	Fails in presence of artificial light intensities

Jones [22]. Let the sequence of face regions be $\{F_1, F_2, \dots F_t\}$

Discriminative response map fitting [23] organizes 68 different landmarks on any face region as shown in Fig. 1. Sudden display of images invokes various emotion responses on the face and it

is reflected as activations in the certain 68 different facial landmarks. Quiroz et al [24] have mapped different emotions to activations in the particular facial landmark areas. The mapping is given in Table 2. A square patch of size m is created for each of set of facial landmarks and this becomes the ROI set. $ROI = \{S(L_1, m), S(L_2, m), \ldots S(L_n,\text{m})\}$

Where S is the square with size m and center point as L_x. The change in the mean intensity values over these ROI regions for all images in sequence $\{F_1, F_2, \ldots F_t\}$ are collected as below

The intensity values in time domain are preprocessed using three temporal filters. De-trending filter is first applied to reduce the slow or non-stationary trend of the signal. Moving average filter is applied to remove random noises. Finite impulse bandpass filter is then applied with cut off frequency of [0.7, 4] Hz. From the preprocessed signals of each ROI region power spectral density (PSD) is calculated. An activation threshold (T) is set by taking the peak value of power spectrum density of ROI regions around all landmarks as

$$T = \frac{\sum_{i=1}^{68} Peak(PSD_i)}{68} \quad (1)$$

In case of ROI is activated, due to image impulse, the peak value of PSD for that ROI region will be more than the threshold. In case ROI is not activated due to image impulse, the peak value of PSD for that ROI region will be less than the threshold. A sample PSD plot for activated ROI region and not activated ROI region is given in Fig. 2.

The left part in Fig. 2 is the activated ROI region, where peak point is above the threshold T. The right part in Fig. 5 is the non activated ROI region where peak point is below the threshold T.

The ROI regions activated are found for the challenge image whose emotion label is already known. To facilitate this, a collection of images are kept in following categories of happiness, sad, fear, angry, surprised, disgusted, awed, appalled and hatred. Image of the particular category is presented as challenge whose activated landmarks be $\{EL_1, El_2, \ldots EL_m\}$ and the activated ROI regions are found in the response. Let the landmarks corresponding to those activation regions be $\{L_1, l_2, \ldots L_n\}$, then a matching score (MS) is calculated as

$$MS = \frac{\sum(EL \cap L)}{\max(m,n)} \quad (2)$$

The challenge image is moved randomly at

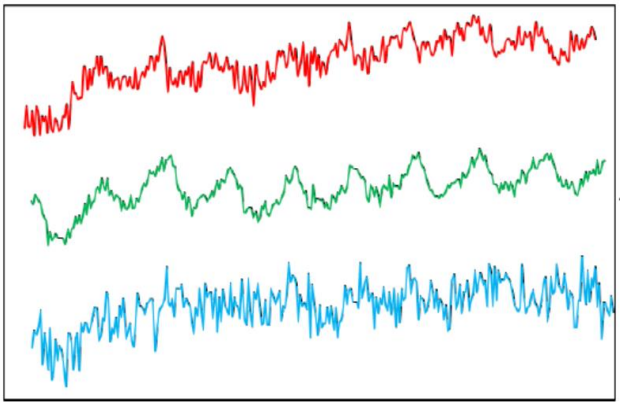

Figure. 1 Intensity plot for the ROI regions

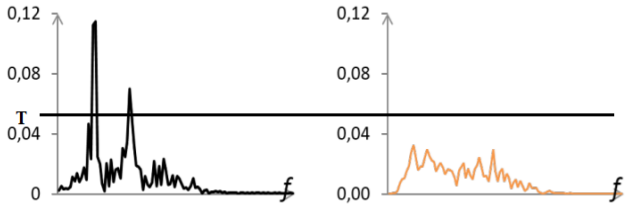

Figure. 2 PSD Plot

positions $\{(x_1, y_1), (x_2, y_2) .. (x_n, y_n)\}$ when face region sequence $\{F_1, F_2, \ldots F_t\}$ are collected. Let the center position of left iris in these face regions be $\{(Ix_1, Iy_1), (Ix_2, Iy_2) .. (Ix_n, Iy_n)\}$. Distance the challenge image has moved over the time period is measured in terms of Euclidean distance as $p = \{p_1, p_2 \ldots . p_{n-1}\}$ etc where

$$p_t = \sqrt{(x_{t+1} - x_t)^2 + (y_{t+1} - y_t)^2} \quad (3)$$

Similarly the distance the center of iris has shifted over the time period is measured in terms of Euclidean distance as $q = \{q_1, q_2 \ldots q_{n-1}\}$ etc where

$$q_t = \sqrt{(Ix_{t+1} - Ix_t)^2 + (Iy_{t+1} - Iy_t)^2} \quad (4)$$

correlation is calculated between p, q as

$$C = \frac{\sum(p_i - \bar{p})(q_i - \bar{q})}{\sqrt{\sum(p_i - \bar{p})^2 + (q_i - \bar{q})^2}} \quad (5)$$

When MS is more than 0.7 or C is more than 0.7 then liveliness is detected and passed to next stage of recognition. In failure of this condition, spoofing is detected, and acquisition process is repeated.

B. Hybrid deep feature based recognition

Local binary pattern (LBP) proposed by Ojala et al [25] is a effective texture pattern descriptor which

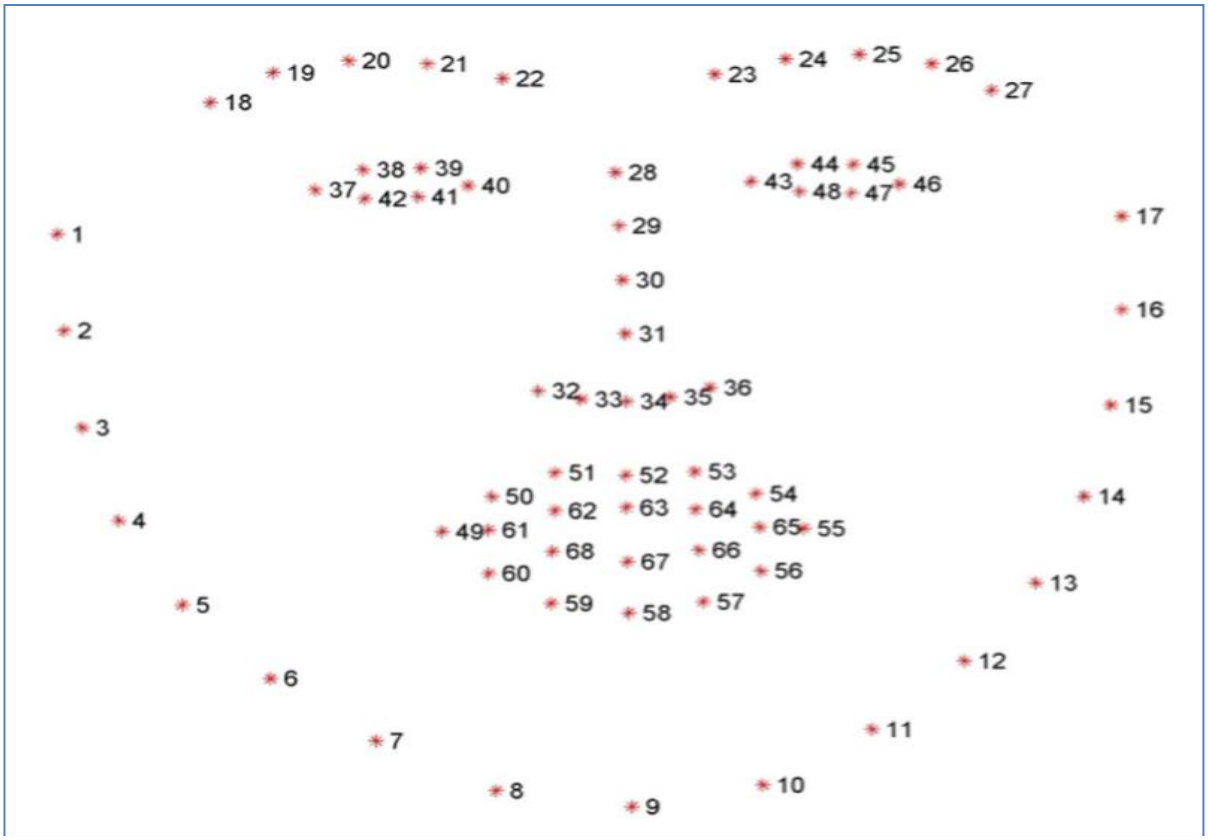

Figure. 3 68 facial landmarks (Courtesy [23])

Table 2. Emotion to facial landmark activation mapping [24]

Emotion	**Facial landmarks activated**
Happiness	12,25
Sad	4,15
Fear	1,4,20,25
Angry	4,7,24
Surprised	1,2,25 ,26
Disgusted	9,10,17
Happily sad	4,6,12,25
Happily surprised	1,2,12,25
Happily disgusted	10.12,25
Sadly fearful	1,4,15,25
Sadly angry	4,7,15
Sadly surprised	1,4,25,26
Awed	1,2,5,25
Appalled	4,9,10
Hatred	4,7,10

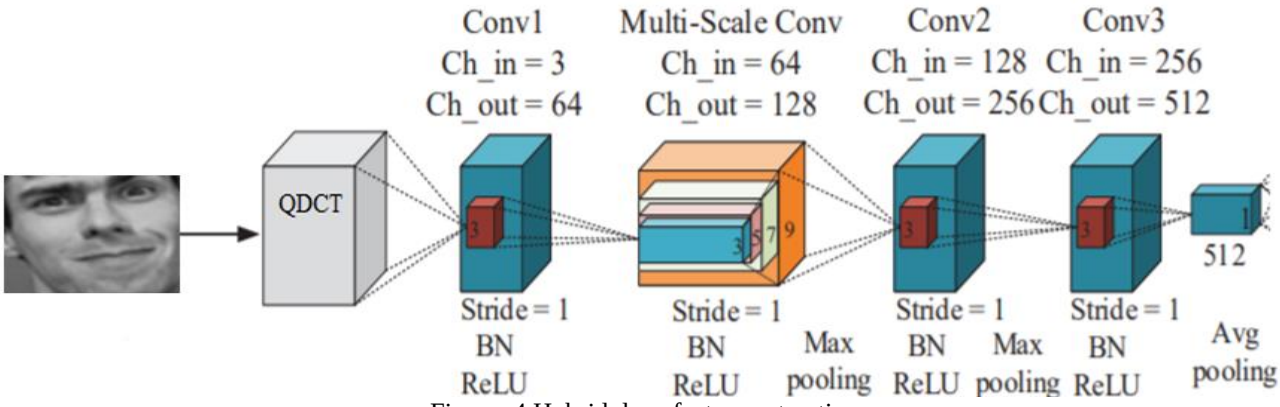

Figure. 4 Hybrid deep feature extraction

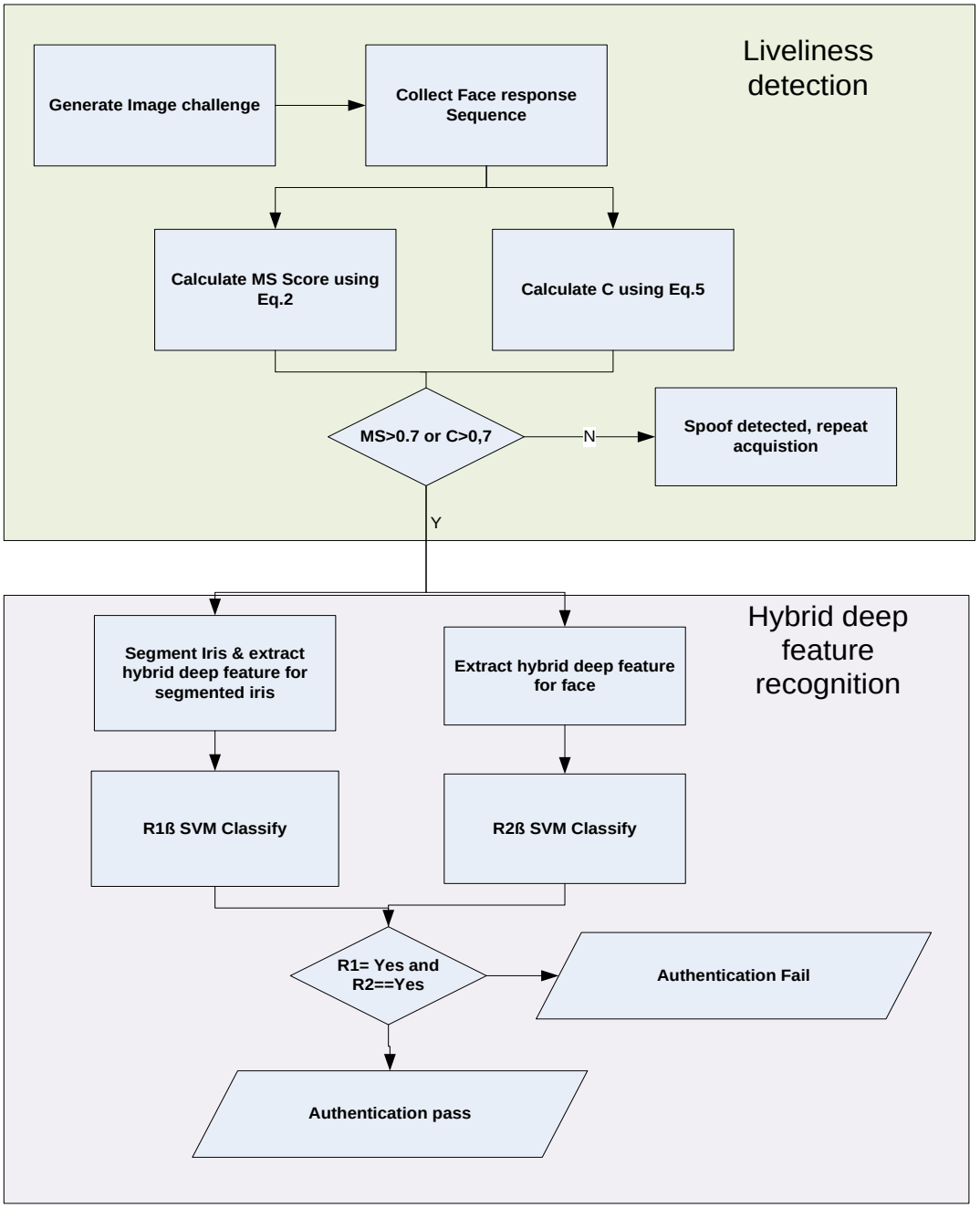

Figure. 5 Process flow

can provide a better representation of local texture pattern of image. Working in a block size of 3×3 , the center pixel is used as threshold for the neighboring pixel. The LBP code of center pixel is generated by encoding the computed threshold into a decimal value. LBP is given as

$$LBP = \sum_{i=0}^{P-1} s(n_i - G_c)2^i \quad (6)$$

$$s(x) = \begin{cases} 1, if\ x > 0 \\ 0, otherwise \end{cases} \quad (7)$$

In the above equation, P is the number of neighborhood pixels, n_i is the i[th] neighboring pixel and c is the center pixel. Souza et al [1] integrated LBP into the first layer of convolutional neural network and referred as LBP-net. The convolutions are done on the LBP transformed pixels instead of original pixels. The method provided better accuracy for face matching in presence of spoofing attacks. In this work, a deviation is made similar to LBPnet architecture by extracting QDCT coefficients from the face or iris regions in first layer and passing the QDCT coefficients to subsequent layer in convolutional neural network.

QDCT for an image $f(x, y)$ is calculated as

$$f(x,y) = A_n^q f(x,y) + \sum_{s=1}^{n}[D_{s,1}^q f(x,y) + D_{s,2}^q f(x,y) + D_{s,3}^q f(x,y)] \quad (8)$$

Where $A_n^q f(x,y)$ is the low frequency band and $D_{s,1}^q f(x,y)$ is the high frequency band of the image. After QDCT is applied on the image a low frequency part, n groups of high frequency parts are obtained. To reduce the dimension of the coefficients, average fusion is done for low frequency sub bands. High frequency sub bands are fused using a fusion rule based on maximum value of energy of coefficients. The average fusion rule for fusing the low frequency bands is given as average of the coefficients pair wise between the Low frequency coefficients of two patch images. The fusion rule for fusing the high frequency sub bands is given as selecting the maximum value of coefficient between the pair wise high frequency sub bands.

The QDCT coefficients are given as input to a frequency domain convolutional neural network (Fig. 2). The coefficients pass through a sequence of ReLU and max pooling layer and a final average pooling layer to provide an output of 1×512 dimension feature vector.

This feature is then used as input for the SVM classifier to recognize person based on face or iris image.

The process flow of the proposed challenge response multimodal biometric recognition is given in Fig. 3. There are two stages, liveliness detection and hybrid deep feature based recognition. In the liveliness stage, challenge image is placed at different locations and face response is collected. From the sequence of face responses, MS is calculated using Eq. (2) and C is calculated using Eq. (5). When MS or C value is more than threshold value of 0.7, recognition stage is trigged else acquisition is repeated. In the recognition stage, hybrid deep feature is extracted from segmented iris and face regions using CNN model based on QDCT given in Fig. 2. The features are classified using respective SVM classifier. The classifiers are trained to classify the face and iris. If both face and iris matches, the user is authentication else the authentication fails.

5. Results

The results are in two sub sections of liveliness detection and recognition.

A. Liveliness detection

The performance of the proposed solution to detect spoofing is tested against OULU-NPU dataset [27]. The dataset has 990 real face videos, 3,960 fake face videos. The performance of the proposed solution is compared against attention based solution proposed by Zheng et al (2021) [8] and spatial gradient solution proposed by Wang et al (2020) [17]. Though the work [8] and [17] were tested for number of datasets, in this work we restrict the comparison to OULU-NPU dataset alone. The results for OULU-NPU dataset for [8] (in Table 6 of [8]) and for [17] (in Table 3 of [17]) were taken. The proposed solution was implemented in python and results were collected for same conditions in [8] and [17].

Table 3. Testing environment

Env1	under random lighting and background.
Env2	random attack media.
Env3	transformation of the attack camera equipment.
Env4	All above three factors combined

Table 4. Env1 results

Env1			
Solution	**APCER %**	**BPCER %**	**ACER %**
Zhen et al (2021)	1.4	1.8	1.6
Wang et al (2020)	2.0	0.0	1.0
Proposed	0.62	0.51	0.32

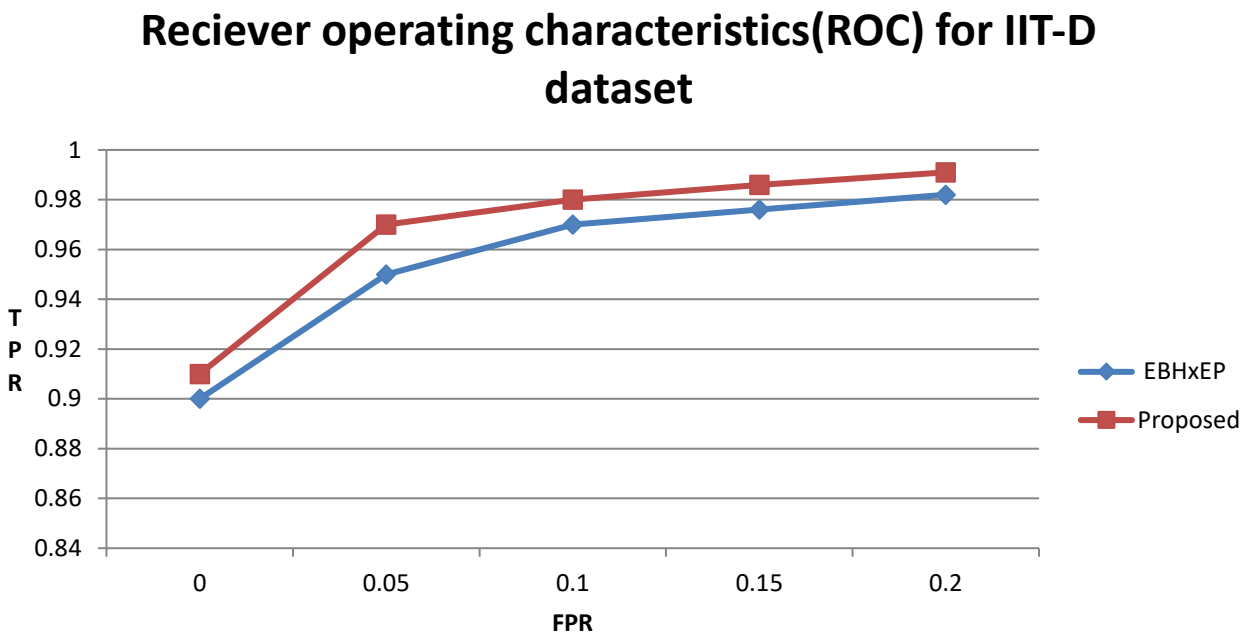

Figure. 2 ROC for IRIS recognition for IIIT-D dataset

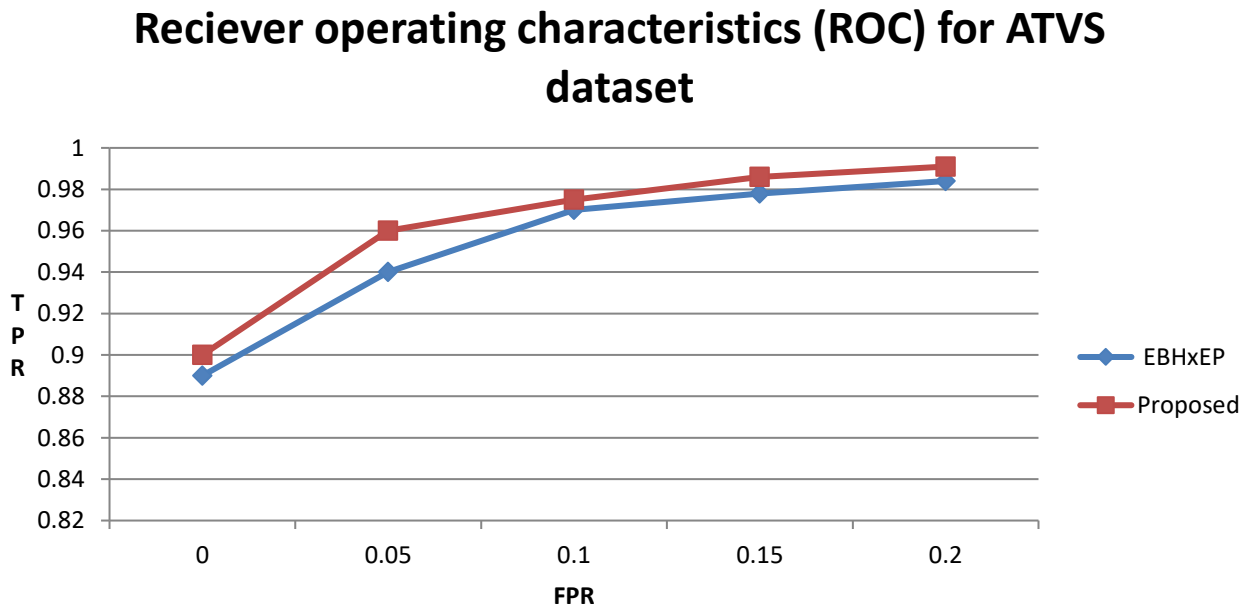

Figure. 3 ROC for IRIS recognition for ATVS dataset

The performance is compared in terms of attack presentation classification error rate (APCER), bona fide presentation classification error rate (BPCER), and average classification error rate (ACER). The lower the values of these error rates, the performance is better.

The performance test is conducted in four environments.

The performance in Env1 are measured and given in Table 4.

The proposed solution has lower error rate in presence of random lighting and background. This is because, the challenge response focused on areas of interest around AU in terms of pixel intensities in proposed solution instead of overall texture patterns used in existing works. The features used in proposed solution were more robust to lighting and background conditions.

The performance in Env2 are measured and given in Table 5.

The proposed solution has lower error rate in presence of random attack media. This is because, pixel intensity features used in this work has no correlation to attack media like screen, mobiles etc. But existing works used texture features which are sensitive to screen reflections.

The performance in Env3 are measured and given in Table 6.

The proposed solution has lower error rate in presence of various camera tuning parameters. This is because, the pixel intensities in AU regions has linear correlation to camera tuning parameters in the proposed solution.

The performance in Env4 are measured and given in Table 7.

The proposed solution has lower error rate in presence of combination of Env1 to Env3. This is due to linear correlation of features to environmental conditions.

From the results, the proposed solution is found to have lower values of error compared to existing works. The proposed solution is more robust to changes in attack pattern and lighting. Use of random challenge has lowered the error in classification between real and fake samples in the proposed solution.

B. Recognition

The face recognition performance of the proposed hybrid deep feature with SVM classifier is measured in terms of accuracy (ACC), false acceptance rate (FAR) and false rejection rate (FRR) for the face images in ORL face dataset [28].

The performance is compared against LBP-net and n-LBPnet proposed in Souza et al [1] for four environment conditions. The performance LBP-net and n-LBPnet was tested against NUAA dataset [34] in [1]. Compared to NUAA dataset, ORL face dataset is rich with acquisitions in different lighting conditions and thus NUAA dataset was used for testing the recognition performance in this work. In addition to proposed solution, LBPnet and n-LBPnet were implemented in python for evaluating the addition to proposed solution, LBPnet and n-LBPnet were implemented in python for evaluating the performance.

Table 5. Env2 results

Env2			
Solution	**APCER %**	**BPCER %**	**ACER %**
Zhen et al (2021)	2.6	0.8	1.7
Wang et al (2020)	2.5	1.3	1.9
Proposed	0.64	0.56	0.41

Table 6. Env3 results

Env3			
Solution	**APCER %**	**BPCER %**	**ACER %**
Zhen et al (2021)	2.0	3.9	2.8
Wang et al (2020)	3.2	2.2	2.7
Proposed	0.71	0.61	0.58

Table 7. Env4 results

Env4			
Solution	**APCER %**	**BPCER %**	**ACER %**
Zhen et al (2021)	4.2	4.6	4.4
Wang et al (2020)	6.7	3.3	5.0
Proposed	1.81	1.61	2.36

Table 8. Face recognition performance in Env1

Env1			
Method	**ACC**	**FAR**	**FRR**
LBPnet	0.976	0.028	0.016
n-LBPnet	0.982	0.019	0.015
Proposed	0.991	0.012	0.013

Table 9. Face recognition performance in Env2

Env2			
Method	**ACC**	**FAR**	**FRR**
LBPnet	0.972	0.029	0.015
n-LBPnet	0.980	0.018	0.016
Proposed	0.992	0.011	0.011

Table 10. Face recognition performance in Env3

Env3			
Method	**ACC**	**FAR**	**FRR**
LBPnet	0.961	0.030	0.015
n-LBPnet	0.972	0.021	0.013
Proposed	0.993	0.011	0.011

Table 11. Face recognition performance in Env4

Env4			
Method	**ACC**	**FAR**	**FRR**
LBPnet	0.951	0.034	0.019
n-LBPnet	0.962	0.027	0.018
Proposed	0.991	0.014	0.015

Table 12. IRIS recognition performance

Database	**IIIT-D**	
Method	**AER**	**ACA**
EBHxEP	3.9	0.981
Proposed	3.2	0.993
Database	**ATVS**	
EBHxEP	1.7	0.984
Proposed	1.3	0.996

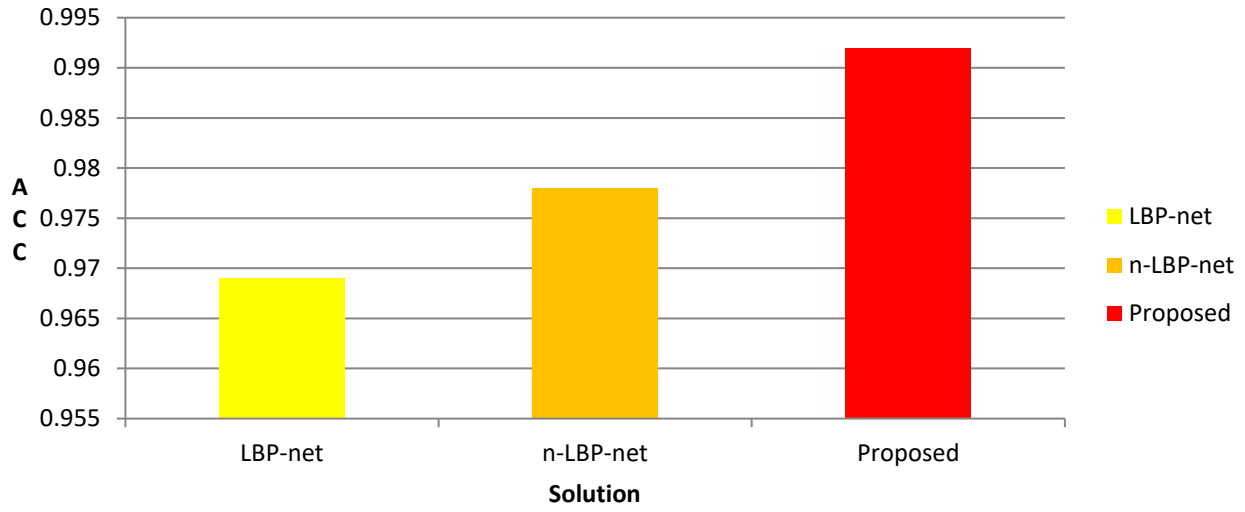

Figure. 4 Comparison of average ACC for face recognition

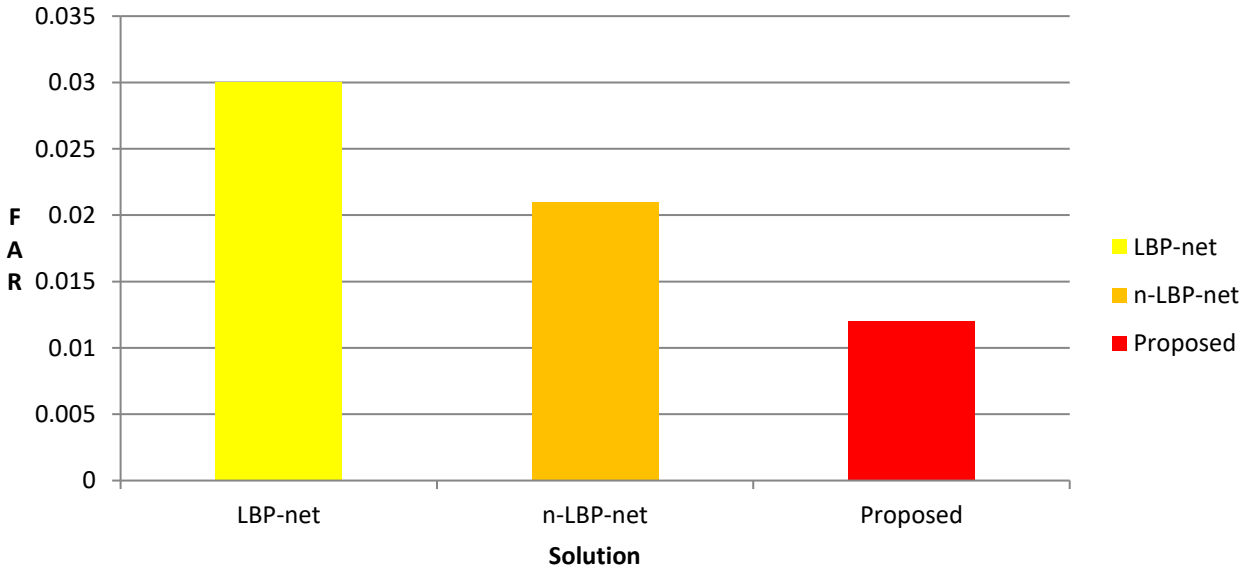

Figure. 5 Comparison of average FAR for face recognition

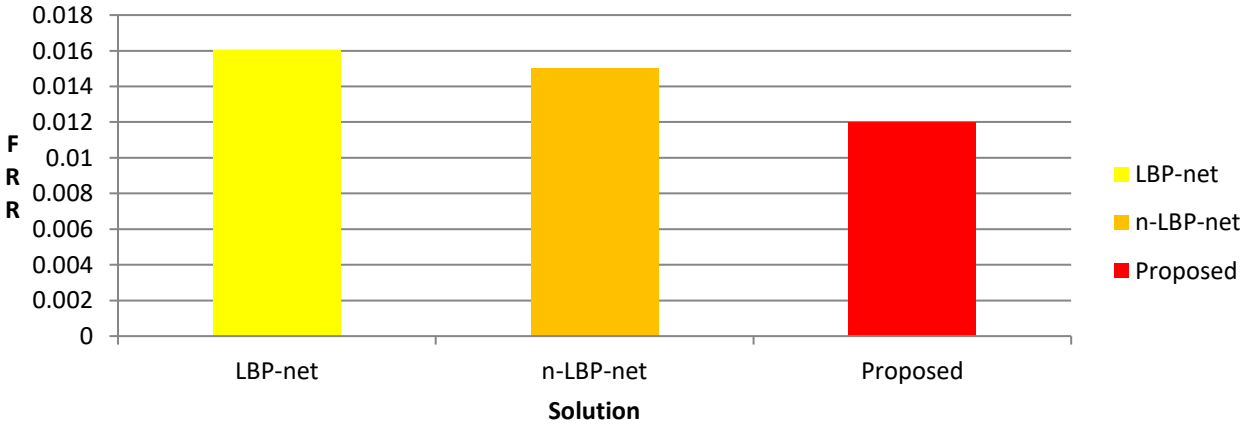

Figure 6. Comparison of average FRR for face recognition

The face recognition performance for Env1 is given in Table 8.

Compared to existing work, the proposed solution has atleast 1 % higher accuracy in presence of random lighting. This is due to QDCT based convolutional features.

The face recognition performance in Env2 is given in Table 9.

Compared to existing work, the proposed solution has atleast 1 % higher accuracy in presence of various attack media.

The face recognition performance in Env3 is given in Table 10.

Compared to existing work, the proposed solution has atleast 2 % higher accuracy in presence of various camera tuning parameters. Effects like zoom, colors had little impact on QDCT based convolutions in proposed solution.

The face recognition performance in Env4 is given in Table 11.

The average ACC across all four environments is given in Fig. 8. The average FAR across all four environments is given in Fig. 9. The average FRR across all four environments is given in Fig. 10.

The proposed solution has 2 % higher accuracy compared to exiting works. The FAR and FRR are low compared to existing works. Use of QDCT coefficients for convolutions has increased the recognition accuracy in the proposed solution. Proposed QDCT fusion has selected the best set of discriminative features in face and this is increased the feature learning ability of the convolutional neural network. Also the proposed solution performance in terms of ACC, FAR and FRR is consistent across all four environments.

The iris recognition performance of proposed hybrid deep feature along with SVM classifier is measured in terms of average error rate (AER) and average classification accuracy (ACA). The performance is tested for IIIT-D CLI database [29] and ATVS dataset [30]. The performance of the proposed solution is compared against EBHxEP proposed by Agarwal et al [18]. The results are given in Table 12.

The AER has reduced in proposed solution compared to EBHxEP and ACA is atleast 1 % higher compared to EBHxEP. The result is consistent for both the datasets. The ROC for IIIT-D dataset is given in Fig. 6 and ROC for ATVS dataset is given in Fig. 7. The ROC provides True positive rate (TPR) for various false positive rates(FPR). From the figures, it can be seen that true positive rate is higher in proposed solution compared to EBHxEP in both datasets and higher positive rate is achieved in IIIT-D dataset.

6. Conclusion

A challenge responsive multi modal biometric authentication using face and iris biometric is proposed in this work. The solution had two stages of liveliness detection and hybrid deep feature based recognition. Liveliness was detected using arbitrary emotion proving images moved randomly over screen and collecting face responses. Hybrid deep features were extracted using a QDCT based deep learning model. The proposed solution allowed spoofs only with error rate of 0.95 % compared to 2.55 % in existing works. Also the face recognition accuracy is atleast 2 % higher and iris recognition accuracy is atleast 1 % higher compared to existing works. The effectiveness of proposed liveliness detection can be tested for different test sets and artificial fakes as part of future work.

Conflicts of interest

The authors declare no conflict of interest.

Author contributions

The paper background work, conceptualization, methodology, dataset collection, implementation, result analysis and comparison, preparing and editing draft, visualization have been done by first author. The supervision, review of work and project administration, has been done by second author.

References

[1] D. Souza, S. Santos, and R. Pires, “Deep texture features for robust face spoofing detection”, *IEEE Trans. Circuits Syst. II Express Briefs*, Vol. 64, No. 12, pp. 1397-1401, 2017.

[2] S. Parveen, S. Ahmad, and H. Abbas, “Face liveness detection using dynamic local ternary pattern (DLTP)”, *Computers,* Vol. 5, No. 2, pp. 10, 2016.

[3] Z. Akhtar and L. Foresti, “Face spoof attack recognition using discriminative image patches”, *J. Electr. Comput. Eng,* Vol. 2016, pp. 1–14, 2016.

[4] J. Zhou, K. Shu, P. Liu, J. Xiang, and S. Xiong, “Face Anti-Spoofing Based on Dynamic Color Texture Analysis Using Local Directional Number Pattern”, In: *Proc of 25th International Conference on Pattern Recognition (ICPR),* Milan, Italy, pp. 4221-4228, 2021.

[5] X. Li, J. Komulainen and G. Zhao, “Generalized face anti-spoofing by detecting pulse from face videos”, In: *Proc of the 23rd International Conference on Pattern Recognition (ICPR),* Cancun, Mexico, pp. 4244–4249, 2016.

[6] R. Hasan, H. Mahmud, and Y. Li, "Face Anti-Spoofing Using Texture-Based Techniques and Filtering Methods", *Journal of Physics: Conference Series,* Vol. 1229, No. 012044, 2019.

[7] R. Cai, R. Wan, and H. Li "Learning Meta Pattern for Face Anti-Spoofing", *IEEE T-IFS,* arXiv: 2110.06753, 2021.

[8] W. Zheng, M. Yue, S. Zhao, and S. Liu, "Attention-Based Spatial-Temporal Multi-Scale Network for Face Anti-Spoofing", *IEEE Transactions on Biometrics, Behavior, and Identity Science*, Vol. 3, No. 3, pp. 296-307, 2021.

[9] X. Song, X. Zhao, and T. Lin, "Face Spoofing Detection by Fusing Binocular Depth and Spatial Pyramid Coding Micro-Texture Features", In: *Proc of IEEE International Conference on Image Processing,* pp. 96-100, 2017.

[10] R. Cai, H. Li, S. Wang, and C. Chen, "DRL-FAS: A Novel Framework Based on Deep Reinforcement Learning for Face Anti-Spoofing", *IEEE Transactions on Information Forensics and Security,* Vol. 16, pp. 937-951, 2020.

[11] Z. Yu, X. Li, X. Niu, J. Shi, and G. Zhao, "Face anti-spoofing with human material perception", *In book: Computer Vision – ECCV,* pp. 557-575, 2020.

[12] X. Tu, H. Zhang, and M. Xie, "Enhance the Motion Cues for Face Anti-Spoofing using CNN-LSTM Architecture", ArXiv, abs/1901.05635,2019.

[13] G. Wang, C. Lan, H. Han, S. Shan, and X. Chen, "Multi-modal face presentation attack detection via spatial and channel attentions", In: *Proc of IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW),* pp. 1584-1590, 2019.

[14] Y. Liu, Y. Tai, J. Li, F. Huang, and R. Ji, "Aurora guard: Real-time face anti-spoofing via light reflection", arXiv:1902.10311, 2019.

[15] C. Chou, "Presentation attack detection based on score level fusion and challenge-response technique", *The Journal of Supercomputing,* Vol. 77, pp. 4681–4697 .2021.

[16] E. Boutellaa, Z. Boulkenafet, and J. Komulainen, "Audiovisual synchrony assessment for replay attack detection in talking face biometrics", *Multimed Tools Appl.,* Vol. 75, No. 9, pp. 5329-5343, 2016.

[17] Z. Wang, Z. Yu, C. Zhao, and Z. Lei, "Deep spatial gradient and temporal depth learning for face anti-spoofing", In: *Proc of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* pp. 5042-5051, 2020.

[18] R. Agarwal, S. Jalal, and V. Arya, "Enhanced Binary Hexagonal Extrema Pattern (EBHXEP) Descriptor for Iris Liveness Detection", *Wireless Personal Communications,* Vol. 115, pp. 2627–2643, 2020.

[19] S. Bhogal, D. Söllinger, and A. Uhl, "Non-reference image quality assessment for biometric presentation attack detection", In: *Proc. of IEEE 5th International Workshop on Biometrics and Forensics,* pp. 1–6, 2021.

[20] A. Fathy and S. Hanaa, "Entropy with local binary patterns for efficient iris liveness detection", *Wireless Personal Communications,* Vol. 102, Issue. 3, pp. 2331-2344, 2018.

[21] T. Nguyen. D. Pham, and W. Lee, "Deep learning-based enhanced presentation attack detection for iris recognition by combining features from local and global regions based on NIR camera", *Sensors,* Vol. 18, No. 8, pp. 2601, 2018.

[22] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," In: *Proc of CVPR,* pp. 511–518, 2001.

[23] A. Asthana, S. Zafeiriou, S. Cheng, and M. Pantic, "Robust discriminative response map fitting with constrained local models," In: *Proc of CVPR,* pp. 3444-3451, 2013.

[24] F. Benitez, R. Srinivasan, and M. Martinez, "EmotioNet: An accurate, real-time algorithm for the automatic annotation of a million facial expressions in the wild", In: *Proc of the IEEE Conference on Computer Vision and Pattern Recognition,* Las Vegas, NV, USA. 26, pp. 5562–5570, 2016.

[25] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray-sclae and rotation invariant texture classification with local binary patterns", *IEEE Transaction on Pattern Analysis and Machine Intelligence,* Vol. 24, No. 7, pp. 971-987, 2002.

[26] C. Galdi, M. Nappi, D. Riccio, and H. Wechsler, "Eye movement analysis for human authentication: a critical survey", *Pattern Recognition Letters,* Vol. 84, pp. 272–283, 2016.

[27] Z. Boulkenafet, J. Komulainen, X. Feng, and A. Hadid, "Oulu-npu: A mobile face presentation attack database with real-world variations", In: *Proc IEEE International Conference on Automatic Face & Gesture Recognition (FG),* pp. 612–618, 2017.

[28] https://cam-orl.co.uk/facedatabase.html

[29] D. Yadav, N. Kohli, and M. Vatsa, "Unraveling the effect of textured contact lenses on iris recognition", *IEEE Transactions on*

Information Forensics and Security, Vol. 9, No. 5, pp. 851–862, 2014.

[30] J. Galbally, J. O. Lopez, and J. O. Garcia, "Iris liveness detection based on quality related features", In: *Proc. of 5th IEEE APR International Conference on Biometrics (ICB)*, pp. 271–276, 2012.

[31] G. Gavisiddappa, S. Mahadevappa, and C. Patil, "Multimodal Biometric Authentication System Using Modified ReliefF Feature Selection and Multi Support Vector Machine", *International Journal of Intelligent Engineering and Systems*, Vol. 13. pp. 1-12, 2020.

[32] G. Bouchiba, R. Tlemsani, and S. Chouraqui, "An Improved Behavioral Biometric System based on Gait and ECG signals", *International Journal of Intelligent Engineering and Systems*, Vol. 12, No. 6, pp. 147-156, 2019.

[33] H. Abed, H. Alsaeedi, and M. Otebolaku, "Palm Vein Identification Based on Hybrid Feature Selection Model", *International Journal of Intelligent Engineering and Systems*, Vol. 14, No. 5, arXiv:2007.16195.2021.

[34] X. Tan, Y. Li, J. Liu, and L. Jiang, "Face liveness detection from a single image with sparse low rank bilinear discriminative model", In: *Proc. Eur. Conf. Comput. Vis*, pp. 504–517, 2010.