Spoofed Facial Presentation Attack Detection by Multivariate Gradient Descriptor in Micro-Expression Region

Dhiman Karmakar^{a,*}, Puja Mukherjee^{a,**}, and Madhura Datta^{b,***}

^aDepartment of Computer Science, Surendranath College, Kolkata, 700009 India

^bUGC-Human Resource Development Centre, University of Calcutta, Kolkata, 700009 India

*e-mail: dhiman.karmakar@gmail.com

**e-mail: puja.uit@gmail.com

***e-mail: mdhrdc@caluniv.ac.in

Abstract—Facial video presentation is a topic of interest in many security systems due to its non-intrusive nature. However, such systems are vulnerable to spoof attacks made by fake face videos and thereby gaining unauthorized access in the system. For a robust biometric system anti spoofing approaches like liveness detection ought to be implemented in order to counter the aforesaid print and replay attacks. This article proposes a novel approach of anti spoofing using Multivariate histogram of oriented gradients descriptor in the auto detected micro expression (μE) regions of human facial videos. Facial μE are very brief, spontaneous facial expressions that highlight the face of humans when they either unconsciously or deliberately conceal an emotion. The work emphasizes the variations in μE in fake and original video representation by a considerable amount and claims such a variance is a tool to combat against presentation attacks. In particular, the method automatically extracts the ROI of major changes in μE using the Multivariate orientation gradients parameter and thus proposes this descriptor as one of the most suitable tools to characterize the liveness. The entire implementation is carried out on the self-created Database for replay attacks. The result obtained is satisfactory and tested statistically significant.

Keywords: facial spoofing, presentation attack, gradient orientation, micro expression, facial control point

DOI: 10.1134/S1054661821020097

1. INTRODUCTION

Biometric systems are becoming integral part of human society with the increasing need of security at various levels. The primary task of a security system is the verification of an individual's identity in order to prevent impostors from intruding and accessing protected resources. General approaches to authenticate a personnel are password matching or ID card verification. However, such mechanism of identity detection can easily be lost, hampered or stolen and thereby undermine the intended security. Incorporating physical and biological traits of human beings in the system has been the trend in modern biometric security. Liveness detection [1] is an active research area among the field of fingerprint and iris recognition [2, 3] in recent years. Some of the other familiar biometrics like gait, hand geometry, retinal texture have also gained interest in spoof detection. The extraction and fusion of different features in multimodal scenario [4] has gradually been popularized in anti-spoofing paradigm. Liu et al. [5] presented the challenges in replay attack detection based on a multimodal face anti-spoofing dataset CASIA-SURF. However, in facial biometric

domain anti spoofing approaches are limited to some extent. Different head positions or low facial resolution [6] increases the challenge in recognizing faces. The *Liveness* may informally be viewed as an act of differentiating the feature space into living and non-living spaces. Here by fake facial representation attempt we mean placement of a personnel's captured video in front of a sensor instead of his live presence in person. The fake video is assumed to be captured and supplied without any further modification (zero effort). It is a challenging issue to determine the trustworthiness of a biometric system security against spoofing [7, 8]. Facial recognition community has been on a constant search for a proper tool to differentiate fake and live faces. The facial μE is a subtle, short, quick (1/3–1/25 s) and involuntary expression shown on the face of humans when they are trying to conceal their actual emotions, specially in high stake situations. They consists of the seven universal emotions: anger, fear, sadness, disgust, happiness, surprise, and contempt. Here, we present a novel approach to recognize the live faces by the help of comparison of multivariate orientation gradient parameters in the facial ROI where variation in μE is observed.

The paper consists of seven sections enlisted below. Section 2 describes the previous related works based on 3D HOG [9], facial μE [10] detection and different

Received August 7, 2020; revised November 28, 2020;
accepted November 30, 2020

anti spoofing techniques. It is worth mentioning that the tools like μE and 3D HOG are used in conjunction with anti spoofing method in the proposed approach. Section 3 illustrates the proposed algorithm in detail. Section 4 demonstrates the self created database used in the experiment and the way training of data is accomplished. Section 5 consists of experimental results and its analysis. Section 6, the concluding section, deals with the limitation and future scope of the proposed method.

2. STATE OF THE ART

Facial spoof attack is a process in which an attacker can subvert a biometric authentication or recognition system by fake facial video presentation. Face recognition systems based on 2D and 3D images could be exposed to spoofing attacks. These attacks are analyzed in terms of its descriptors and classifiers. Here, anti spoofing is primarily a liveness detection mechanism that aims at detecting physiological signs of human beings such as eye blinking, facial expression changes, mouth movements, etc. Spoofing detection approaches have been categorized into various groups by the researchers. Analysis based on Frequency and Texture is one among them. The texture information taken as 2D images tend to suffer from the loss of texture information compared to the 3D images. In many of the feature extraction schemes, frequency and texture-based fusion are observed in order to form the feature vector. For instance [11] applies 2D DFT to segregate facial image into several groups of concentric rings representing corresponding region in the frequency band. Thereafter the feature vector is acquired by combining the average energy values of all the concentric rings. The Variable Focusing approach [12] suffers from the constraint of relying on Depth of Field (DoF), range between the nearest and farthest objects in a given focus. To increase the liveness detection performance, the authors have increased out focusing effect for which the DoF should be contracted.

Analysis based on Eye Blinking and Eye Movement have also been gaining researchers' interest. An interesting Lie detection algorithm [13] drew its conclusion by observing the changes in pupil dilated using circular Hough transform and increments in the number of eye blinking through frame difference method. The blinking-based approach for liveness detection using Conditional Random Fields (CRFs) [14] to model blinking activities, is current trend for accommodating long-range dependencies on the observation sequence. Often a comparison of CRF model with a discriminating model like AdaBoost and a generative model like HMM is observed in articles. Blinking activity is an action represented by the image sequence consisting of images with close and non-close state. Based on this feature, Pedone and Heikkila [15] illustrate eye region probing method to identify liveness of face. To decrease the effect of illumination experiment

is carried out in Self-Quotient Images (SQI). The 3D Face Shape based analysis [16] in a biometric system differentiates real face from a photograph. It can be implemented as an anti-spoofing tool, coupled with 2D face recognition systems and/or be integrated with a 3D face recognition system for an early detection of spoofing attacks. Image Distortion Analysis (IDA) proposed by Jain et al. [17] seems to be an effective tool in anti-spoofing mechanisms. Bharadwaj et al. [18] proposed a novel approach of using Euclerian motion magnification for the enhancement of facial features in face video spoofing detection. Kollreider et al. [19] have presented a face detection technique which includes mouth localization using motion analysis to confirm liveness. The wavelet decomposition-based face liveness recognition system [20] possesses good discrimination capability between live and fake faces. The voting-based liveness detection system [21] showed promising results compared to existing systems by discriminating image patches of regions that are salient, instrumental, and class-specific.

The μE are more difficult to analyse since they last only $1/3-1/25$ s and involve spatial variations that can hardly be detected with naked eyes. Facial feature extraction plays an important role to recognize the facial μE . The facial features can be classified into couple of major types, geometric and appearance based. Geometric features measure the displacement of facial components such as eyes while appearance based features determine the texture changes on a face when an action such as smiling is performed. Shreva et al. [22, 23] computed the strain magnitude of optical flow to spot μE from macro-expressions by observing the interval flow in a given threshold. Polikovskiy et al. [24] extracted 3D-gradient descriptors from the widely used face region (FACS), using their own 200 fps high-speed camera-recorded database. Park and Kim [25] proposed the detection of subtle facial expression using motion magnification. An interesting approach of combining Local Gradient Pattern with 2D subspace in order to recognize faces can be found in [26]. Chan et al. [27] proposed a face liveness detection method for 2D spoofing attack using flash technique. It reduces the influence of environmental factors with low installation cost and less user collaboration. Surinta et al. [28] claimed that, in gender recognition domain, the Histogram of Oriented Gradient (HOG) descriptor outperforms the scale-invariant feature transform (SIFT) descriptor when combined with the support vector machine (SVM). A face recognition system is designed based on the entropy of each face region in HOG descriptor [29].

Our work emphasizes the selection of multivariate HOG descriptors on human face regions where variations in μE is observed. The multivariate histogram representing gradient descriptor feature from now onward will be abbreviated as MVH.

3. PROPOSED APPROACH

3.1. Method Outline

Below we summarize a brief outline of the entire procedure proposed through a self-explanatory flow graph in Fig. 1. The detailed algorithm is discussed in the next section. Now onward the yellow marker stick to each diagram of Fig. 1 will be termed as St. (Stands for sticker).

Readers are advised not to get confused about the three terms T-face, T-grid, and T-blocks used extensively in this algorithm. The first one signifies the auto-scissored T-like portion of frontal face containing the major control points. The second one specifies the grid like structure of varying number of grids imposed over the T-face image and the third one denotes a cluster of rectangular grids in the T-grids where majority of variations in μE is observed.

At first the facial video representation is separated into frames and fed into the system (St.a). In (St.b) exit mode (state of removal of face from camera focus) or entry mode frames are detected and deleted. Next the major feature points of frontal face is detected (St.c) and consequently a T-shaped feature-rich portion of face containing those control points is cropped automatically (St.d) and passed into the next phase (St.e). The pre-processing phase ends at this juncture.

To identify the region of our interest (ROI) where the major changes in μE are occurring, the T-face is divided into a grid like structure (St.e). Number of such grids are not fixed and is subject to increase to pinpoint the ROI more accurately. Each of the rectangular region within the T-face grid is termed as T-block.

In (St.f) we visualize MVH for a particular T-block. Note that the histogram shown in (St.f) is merely for understanding the procedural abstraction and is not at all intended to reflect our experimental outcome as MVH is virtually impossible to visualize due to its higher number of dimensions.

MVH used as a feature is applied for each T-block to identify the ROI of major μE (see the details of MVH in step 3.2.4 of the algorithm). So identified variations in μE is depicted in (St.g) for already trained fake and real presentations. In (St.i) θ difference (difference of angular orientation between same ROI of two different T-frames exhibiting maximum variance) is computed and estimated with a threshold value. Finally, (St.j) based on this θ difference the ultimate conclusion, namely whether its a spoofed or live presentation is drawn.

3.2. Proposed Algorithm

3.2.1. Video acquisition. In the beginning of our method a facial video of duration not more than 7 s is captured. The video is assumed to be a frontal human face and the rest of this algorithm is designed based on

Fig. 1. Illustration of procedural abstraction.

this assumption. Note that recognizing or authenticating a personnel is not the motto of the procedure, rather to detect the malicious intent of posing a frontal face video instead of originally standing in front of the capturing camera. The approach may be considered as an essential measure prior authenticating a personnel. The experiment is carried out for thirty different individuals (as described in the database section). When an individual P places his face in front of the camera, as per our assumption, his original video $v(P)$ is captured. In some later stages, his captured video $v(P)$ as a fake facial presentation is placed in front of the camera and the video $v(v(P))$ is captured and stored in the database. The objective of this algorithm is to conclude that P is real by analyzing the input $v(P)$ and that $v(P)$ is fake by validating the input $v(v(P))$. In the overall process illumination condition is assumed to be normal and invariable for each person.

3.2.2. Video normalization. The captured video is divided into k frames in different time instances like $t_1, t_2, \dots, t_i, t_{i+1}, \dots, t_j, t_{j+1}, \dots, t_k$. It is trivial that during the entry mode (time difference, $t_q - t_p$ between placing the frontal face within camera focus, say t_q and entering into capturing zone of the camera, say t_p , where $t_q > t_p$) and exit mode (time difference, say $t_s - t_r$, between withdrawing frontal pose from the camera

focus, say at time t_r and exiting from the capturing zone, say at time t_s , where $t_s > t_r > t_q > t_p$) frontal face view won't be taken into consideration. We assume $t_1, t_2, \dots, t_{i-1}$ signifies the entry time and $t_{j+1}, t_{j+2}, \dots, t_k$ the exit time and thereby omit those frames retaining $t_i, t_{i+1}, \dots, t_j$ as the normalized frames. These normalized frames are re-mapped as $t_1, t_2, \dots, t_n$ assuming $n = j - i + 1$.

3.2.3. Formation of T-face. This portion of the algorithm is entirely based on the article [30] which concludes that the major variance region of a frontal human face is situated in the T shaped zone covering two eyes, nose, and mouth as shown in the Fig. 1. (St.D) Incidentally, it has been proven that the region of interest covering the major human facial μE is also situated within the T-face. We may term the process "auto-scissor" as it geometrically detects the major control points of a frontal face frame and automatically crops the so called T-region. The auto-scissor is applied on the frames $t_1, t_2, \dots, t_n$ to obtain the T-faces $T_1, T_2, \dots, T_n$, respectively. Here t_i signifies frame obtained in the i th time stamp. Since the major region of interests for μE fall in the T-region, our further experimental steps are confined and carried out on those T s.

3.2.4. Formation of multivariate gradient histogram descriptor. This step is applied for each of the T_i s. Each pixel in the $R-G-B$ plane is viewed as a 3D point and the so formed cube can be viewed as three matrices viz. R, G , and B holding red, green and blue intensity values respectively. Now a 8×8 sliding window is rotated all over R to obtain magnitude and orientation gradient as $r_m = \sqrt{\delta x^2 + \delta y^2}$ and $r_\theta = \tan^{-1} \left(\frac{\delta y}{\delta x} \right)$ for each pixel (x, y) , respectively, where $\delta x = |(x-1, y) - (x+1, y)|$ and $\delta y = |(x, y-1) - (x, y+1)|$. Likewise g_m , g_θ and b_m , b_θ are calculated for G and B , respectively. Now a MVH mh (histogram cube) is formed where three axes are r_θ , g_θ , and b_θ . Here, $mh(x, y, z)$ denotes cumulative magnitude of all the pixels having gradient direction w , x and y in terms of angle in r_θ , g_θ , and b_θ orientation respectively. Note that as in traditional HOG we distribute the magnitude values in separate bins. Number of bins for each axes is kept identical throughout though the range of angles may vary. For example, let number of bins is kept fixed at 9. Again, let us suppose range of r_θ and g_θ varies from 0° to 180° and 0° to 162° , respectively. Then each bin of r_θ has a capacity of 20° and that of g_θ has a capacity of 18° .

3.2.5. Detection of two T-region and their orientation. Here we find out two T s having maximum variance in gradient vectors retrieved from mh . Ideally these two T s will be responsible for the variance in facial μE . If these two vectors are $\hat{x}$ and $\hat{y}$, both of

length n then we may generate the angle θ between them as $\theta_i = \cos^{-1} \frac{y_i - x_i}{r \sin \theta_{i-1} \sin \theta_{i-2}, \dots, \sin \theta_n}$. A pseudo code to generate θ is depicted below.

```

1:  $r \leftarrow \sqrt{\sum_{i=1}^n (y_i - x_i)^2}$ 
2: for  $i \leftarrow 1$  to  $n$  do
3:    $prod \leftarrow 1$ 
4:   for  $j \leftarrow i - 1$  to 1 by  $-1$  do
5:      $prod \leftarrow prod \times \sin \theta_j$ 
6:   end for
7:    $\theta_i \leftarrow \cos^{-1} \frac{y_i - x_i}{r \times prod}$ 
8: end for

```

3.2.6. Real or fake facial video verdict. A threshold ϵ is estimated by calculating θ s for all the real face posing in the training phase. In the test phase of our experiment θ is again calculated for an undetected test video and if θ is found to be more than threshold ϵ it is considered as a real face and otherwise a spoofed frontal face video.

3.3. Mathematical Intuition

The big question trivially arises into reader's mind is "Why is the θ greater in live faces"? Here, θ depicts difference in the change of orientation in terms of angle between two frames of facial presentation. These two frames are responsible for maximum changes in μE . More specifically, let frame i and j be the T-faces auto-scissored respectively from frame f_i and f_j where maximum variance in gradient orientation is observed in frame f_i and f_j . Let L_i , F_i and U_i represents live, fake and unknown facial presentation type of i th T-face respectively. Let also, θ_{Xij} signifies changes in gradient angles owing to the changes in μE in i and j for X -type facial presentation, where X can be L , F or U . Now let, $\theta_{Lij} = |L_i - L_j|$, $\theta_{Fij} = |F_i - F_j|$, $\theta_{Uij} = |U_i - U_j|$.

It is found that $\theta_{Lij} \gg \theta_{Fij}$ and therefore for a certain angular threshold value T , if $\theta_{Uj} > T$ then U is considered real otherwise fake presentation.

Readers may naturally wonder about such a difference in the θ value of fake video and live pose. Why is the θ difference more in live faces? The query is trivial and the answer is twofold. Firstly, it has been a proven fact that live faces have higher reflectance and illuminance than fake videos and hence possesses a greater θ value.

Secondly, prominence of change in μE is more vividly observed in real face due to the following reason.

Fig. 2. Fact of prominence in change of μE . Here E and E' signify micro expressions before and after the occurred changes respectively.

Let, Figs. 2a–2d respectively represents a μE in a particular T-block t for live face, change in μE for the same live face in the same T-block t in some later instance of time, μE in fake video for t and μE changes for fake video in subsequent frames for t . Now, any video, even with top quality of clarity must not be free of grains (denoted by X in Fig. 2). Now θ change in Figs. 2a, 2b is definitely more significant than that of Figs. 2c, 2d due to the following reason. Let the dots (D) denote original μE edges and cross (X) the grains present in the video. Let dots and crosses in Figs. 2a–2d are represented by D_a, D_b, D_c, D_d and X_a, X_b, X_c, X_d , respectively. Now the θ change in μE is measured by the difference in angle between $\theta_1 = (D_a \cup X_a)$ with $(D_b \cup X_b)$ for real faces and $\theta_2 = (D_c \cup X_c)$ with $(D_d \cup X_d)$ for fake faces. As $X_a \ll X_c$ and $X_b \ll X_d$ the conclusion comes out trivially $\theta_1 \gg \theta_2$.

It is worth mentioning that change in μE from frame to frame is a continuous, gradual and smooth method and any abrupt changes (spikes) should be considered as noise or outlier or malicious fake attempt. Therefore a terminating condition, for a very high θ difference H “if $\theta_U = H \gg T$ then reject as Fake attempt” has been imposed in our algorithm.

Selection procedure of global threshold in our algorithm is quiet straight forward. During the training phase a minimum local threshold t_i is adjusted by *fine tuning* and is applied to each class i for a proper identification of spoofing. The aforementioned threshold $T = \max(t_i), \forall i$ is applied globally during the testing phase. By the term *fine tuning* we mean, during the train period same frames undergo experimentation with different threshold values unless the optimum result is obtained and henceforth the local threshold is fixed. However, in the testing period, different frames undergo experimentation with same threshold (global) value T .

4. DATABASE DESCRIPTION AND TECHNICALITIES

For this particular experiment we have created our own video database of facial presentation. Facial video

of thirty number of individuals are captured in different indoor environment in three different sessions. Each video duration is approximately of six to nine seconds including the entry and exit mode, as mentioned in the algorithm in section 3.2.1. Distance of face object from camera is kept 70 to 90 cm on average. The camera has an image resolution of 4616×3464 pixels with optical image stabilization. It has 20 MP resolution along with LED flash and high dynamic range mode (HDR). It is equipped by a physical aperture of F2.0 with exposure compensation and supports video recording of 1920×1080 at 30 fps (frame per second).

Some of the frames extracted from a specific video presentation, as an instance, is shown in Fig. 3. The database thus constructed and stored is considered as live or real facial presentation. To construct the fake video database all the so stored videos are again placed in front of the camera and the videos are captured. Thereby thirty different classes for real and fake video representation representing each personnel are created. During the experiment, videos are considered to be illumination invariant. It is also worth mentioning that the movement of eyes or head within a video, if any, comes out naturally and is not at all instructed to perform, as μE changes is implicit in nature and need not be explicitly expressed.

5. RESULTS AND EXPERIMENTAL ANALYSIS

As mentioned earlier, 3 sets of presentation videos for an individual is maintained for each of the 30 different classes. Trivially, including the so formed fake videos size of the database get doubled containing a total of $2 \times 3 \times 30$ facial presentations.

For the purpose of validation, the so formed T-regions is divided into number of variable blocks as shown in Fig. 4. A block-wise variance of subsequent T-frames are computed to identify the blocks responsible for maximum variance and hence the μE changes.

The result obtained is shown in Table 1. As shown in Fig. 4 the T-face based on different ROI is being divided into 8×3 grid. We assume these 24 grids are represented as $(0, 0)$ to $(7, 0)$ in matrix notation form.

Fig. 3. Different frames of one facial presentation.

Fig. 4. Variation in number of grids to focus on ROI of μE .

Readers are requested to do a one to one mapping based on grid numbers with Fig. 4 and Table 1.

The Table 1 depicts the changes in gradient orientation with the changes in μE in each of the grids. Since in Fig. 4 the two different frames Figs. 4a, 4b having maximum variance in orientation produces maximum changes in the eye region (especially, eye brow and eye lids) the major percentage changes in matrix (Table 1) encompassing the eye region accordingly. Figures 4b, 4e show the vastly affected areas producing the aforesaid changes. for the purpose of maintaining higher accuracy in result and pinpoint the zone

of μE changes, the number of grids in the frames are subsequently increased. As shown in Figs. 4c, 4d number of grids are increased to 5×5 and the grids producing maximum changes in orientation could be observed in Figs. 4g, 4h, accordingly.

In Fig. 5 readers may visualize the difference in various gradient parameters with respect to a specific T-block. The figure consists of three rows and five columns. The first column indicates the different T-block chosen in this case viz. the difference in μE while eye movement. The 2nd, 3rd, 4th, and 5th column respectively represents horizontal gradients, vertical gradients, magnitude of gradient and angle of gradient corresponding to the T-blocks.

The MVH responsible for θ change due to the change in μE in a T-block is difficult to visualize for its higher dimension order. Although, we have made an attempt of visualization through scatter plots and stem plots. Number of bins in MVH is chosen to be 9, each capable of holding 20° angles. The x, y , and z axis in Figs. 6 and 7 represent angular orientations in the bins responsible for containing the angular orientation of Red, Green, and Blue components of a T-block. A point $P = (r, g, b)$ in Fig. 4 indicates the existence of a point P in the T-block having angular orientation value of r°, g° , and b° in the r_θ, g_θ , and b_θ plane respec-

Table 1. Orientation changes in different μE ROI

Grid	Col 0	Col 1	Col 2
Row 0	68%	2%	89%
Row 1	91%	5%	92%
Row 2	87%	7%	94%
Row 3	88%	25%	90%
Row 4	7%	7%	8%
Row 5	4%	5%	6%
Row 6	12%	8%	15%
Row 7	32%	28%	42%

Fig. 5. Magnitude and direction.

tively. However, the magnitude of this point is not possible to depict in the figure as the number of dimensions exceeds three.

In Fig. 7 stem plots of MVH is designed. Figures 7a, 7b represent the MVH of the same T-block in different frames. The higher and lower value of magnitude are represented by blue and red stems respectively. In Fig. 7c only the higher magnitude values indicating the difference in frame Figs. 7a, 7b is plotted. We are interested in the corresponding angular values in different dimensions of these higher magnitude values. Imposing threshold on these angles is necessary in order to conclude any decision regarding liveness.

The experiment is carried out n number of times changing the count of training and test set as well as combination of classes. Two ROC curves obtained by the experiment respectively for the closed set (training and test class ranges from 1 to 20) and open set (training class 10 to 30 and test class 1 to 10) are depicted in Fig. 8 ROC1 and Fig. 8 ROC2.

Among the entire sets of three presentations, two are used for training purpose and remaining one for testing in all the available 30 classes. The global threshold computed in the training phase is applied in testing. Table 2, with a global threshold $T = 3.50$ depicts arbitrarily 5 instances of live and fake presentations each.

Here columns 1 to 6, respectively, signifies the class (out of 30) in which the person belongs to, the specific set within a class, whether originally its a real or fake presentation, experimental outcome of the measured changes in angular orientation (in terms of degree) of

Fig. 6. Scatterplot of 3D histogram.

μE , the final decision made by the proposed method and lastly whether the decision taken is correct or wrong. For example a tuple (25, 3, Y, 3.97, Y, Y) in Table 2 indicates the 3rd set presentation belonging to 25th class yields a change of 3.97° in μE and the experiment correctly identifies it as areal video. An instance of false acceptance is shown in the penultimate row of Table 2.

6. CONCLUSIONS: LIMITATIONS AND FUTURE CHALLENGES

The novel approach depicted here analyses the presence of live features in facial videos using multi-variate orientation gradients parameter in the μE regions of auto scissored T-faces. The efficacy of the so used feature descriptor has been proven empirically. Here, we intend to use motion magnification in order to accentuate the magnitude of the facial movement

Table 2. Decision made by the proposed method on a random sample of database

Class	Set	Real?	μE change	Verdict real? (Spike)	Correct?
10	3	Y	5.73	Y	Y
15	3	Y	5.78	Y	Y
20	3	Y	4.92	Y	Y
25	3	Y	3.97	Y	Y
30	3	Y	3.94	Y	Y
10	3	N	2.75	N	Y
15	3	N	2.43	N	Y
20	3	N	8.47	N	Y
25	3	N	3.93	Y	N (FA)
30	3	N	2.71	N	Y

Fig. 7. Live and fake presentation stem plot.

Fig. 8. ROC.

during a micro-expression change. The video database is constructed in highly controlled environments. Artificial lighting condition is used and the subjects are restrained to free movement of heads in order to keep a near-frontal pose etc. The algorithm may be modified to detect μE regions of face even from non frontal angles. The 3D facial structure may be taken into account or synthetically generated for this modified setting. The method described here is strongly

threshold dependent and result may drastically vary for a wrong selection of threshold. Any threshold independent approach may be used in conjunction with the proposed approach to verify its stability and robustness.

It is worth mentioning that the area (different grids in Fig. 4) where major changes in μE is observed, though certainly fall within the T-regions, never guarantee any specific or constant cluster location of grids.

That is from time to time interested location of grid may vary. Therefore, to detect variations in μE within some specific ROI, for example, in the eye-corner, the algorithms need to be alleviated to concentrate on particular grids. We have used the well-known t-test method and concludes that the result obtained is statistically significant. The algorithm, in case of fake video detection assumes no further processing (zero effort) or addition of artefacts in the fake facial video once it is captured. However a stronger attacker may challenge the system by further processing of fake presentation.

The proposed liveness detection algorithm, in future, may be integrated into a complete μE analysis framework which not only detects the fake videos but also capable of recognizing the type of μE occurring in the human face. In COVID-19 pandemic scenario masked faces are required to be authenticated. The algorithm, due to its capability of detecting μE in the eye region (not covered by a mask) may be experimented as a facial anti spoofing mechanism in such circumstances.

ACKNOWLEDGMENTS

We sincerely thank Mr. Abhijit Basu, M.Tech student of Jadavpur University, Kolkata, India for his effort to construct our experimental database.

FUNDING

The research work is carried out without any financial grant from any organization/institute. However we would like to sincerely thank Mr. Abhijit Basu, M. Tech student of Jadavpur University, Kolkata, India for his effort to construct our experimental database.

CONFLICT OF INTERESTS

The authors declare that they have no conflicts of interest.

COMPLIANCE WITH ETHICAL STANDARDS

This article does not contain any studies involving animals performed by any of the authors.

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Photographs and video of thirty individuals were captured in order to construct the experimental database. Informed consent was obtained from all individual participants involved in the study.

REFERENCES

1. G. Pan, Z. Wu, and L. Sun, "Liveness detection for face recognition," in *Recent Advances in Face Recognition*, Ed. by K. Delac, M. Grgic, and M. Stewart Bartlett (I-Tech, Vienna, 2008), pp. 236–252.
2. S. Sirohey, A. Rosenfeld, and Z. Duric, "A method of detecting and tracking irises and eyelids in video," *Pattern Recognit.* **35** (6), 1389–1401 (2002).
3. G. Odínokikh, Iu. Efimov, I. Solomatin, M. Korobkin, and I. Matveev, "Iris anti-spoofing solution for mobile biometric applications," *Pattern Recognit. Image Anal.* **28**, 670–675 (2018).
4. M. Sivarathinabala, S. Abirami, and M. Deivamani, "A smart security system using multimodal features from videos," *Pattern Recognit. Image Anal.* **29**, 89–98 (2019).
5. Ajian Liu, Jun Wan, Sergio Escalera, Hugo Jair Escalante, Zichang Tan, Qi Yuan, Kai Wang, Chi Lin, Guodong Guo, Isabelle Guyon, and Stan Z. Li, "Multi-modal face anti-spoofing attack detection challenge," in *CVPR2019* (IEEE, 2019).
6. Naser Zaeri, "Thermal face recognition under spatial variation conditions," *Pattern Recognit. Image Anal.* **30**, 108–124 (2020).
7. J. Galbally, S. Marcel, and J. Fierrez, "Biometric anti-spoofing methods: A survey in face recognition," *IEEE Access* **2**, 1530–1552 (2014).
8. G. Pan, Z. Wu, L. Sun, and S. Lao, "Eyeblink-based anti-spoofing in face recognition from a generic web-camera," in *2007 IEEE 11th International Conference on Computer Vision* (2007).
9. N. Buch, J. Orwell, and S. A. Velastin, "3D extended histogram of oriented gradients (3Dhog) for classification of road users in urban scenes," in *British Machine Vision Conference* (2009).
10. R. Danescu, D. Borza, and R. Itu, "Detecting micro-expressions in real time using high-speed video sequences," in *Intelligent Video Surveillance* (IntechOpen, 2018).
11. G. Kim, S. Eum, J. K. Suhr, D. I. Kim, K. R. Park, and J. Kim, "Face liveness detection based on texture and frequency analyses," in *2012 5th IAPR International Conference on Biometrics (ICB)* (2012), pp. 67–72.
12. Sooyeon Kim, Sunjin Yu, Kwangtaek Kim, and Yuseok Ban, "Face liveness detection using variable focusing," in *2013 International Conference on Biometrics (ICB)* (IEEE, 2013).
13. Respatyadi Hari Nugroho, Muhammad Nasrun, and Casi Setianingsih, "Lie detector with pupil dilation and eye blinks using Hough transform and frame difference method with fuzzy logic," in *2017 International Conference on Control, Electronics, Renewable Energy, and Communications (ICCREC)* (IEEE, 2017).
14. Lin Sun, Gang Pan, Zhaohui Wu, and Shihong Lao, "Blinking-based live face detection using conditional random fields," in *ICB 2007: Advances in Biometrics* (2007), pp. 252–260.
15. M. Pedone and J. Heikkilä, "Local phase quantization descriptors for blur robust and illumination invariant recognition of color textures," in *ICPR* (2012), pp. 2476–2479.
16. A. Lagorio, M. Tistarelli, and M. Cadoni, "Liveness detection based on 3D face shape analysis," in *2013 International Workshop on Biometrics and Forensics (IWBF)* (IEEE, 2013).

17. D. Wen, H. Han, and A. K. Jain, "Face spoof detection with image distortion analysis," *IEEE Trans. Inf. Forensics Secur.* **10** (4), 746–761 (2015).
18. S. Bharadwaj, T. Dhamecha, M. Vatsa, and R. Singh, "Computationally efficient face spoofing detection with motion magnification," in *2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops* (2013), pp. 105–110.
19. K. Kollreider, H. Fronthaler, M. I. Faraj, and J. Bigun, "Real-time face detection and motion analysis with application in liveness assessment," *IEEE Trans. Inf. Forensics Secur.* **2** (3), 548–558 (2007).
20. Md Moniruzzaman and Mohammad S. Alam, "Wavelet decomposition-based efficient face liveness detection," *Proc. SPIE* **9845** (2016).
21. Zahid Akhtar and Gian Luca Foresti, "Face spoof attack recognition using discriminative image patches," *J. Electr. Comput. Eng.* **2016**, 4721849 (2016).
22. M. Shreve, S. Godavarthy, V. Manohar, D. Goldgof, and S. Sarkar, "Towards macro- and micro-expression spotting in video using strain patterns," in *WACV* (2009), pp. 1–6.
23. M. Shreve, S. Godavarthy, D. Goldgof, and S. Sarkar, "Macro- and micro-expression spotting in long videos using spatio-temporal strain," in *2011 IEEE International Conference on Automatic Face & Gesture Recognition (FG)* (2011), pp. 51–56.
24. S. Polikovsky, Y. Kameda, and Y. Ohta, "Facial micro-expressions recognition using high speed camera and 3D-gradient descriptor," in *ICDP* (2009), pp. 1–6.
25. S. Park and D. Kim, "Subtle facial expression recognition using motion magnification," *Pattern Recognit. Lett.* **30** (7), 708–716 (2009).
26. A. Dahmouni, N. Aharrane, K. El Moutaouakil, and K. Satori, "A face recognition based biometric solution in education," *Pattern Recognit. Image Anal.* **28** (4), 758–770 (2018).
27. Patrick P. K. Chan, Weiwen Liu, Danni Chen, Daniel S. Yeung, Fei Zhang, Xizhao Wang, and Chien-Chang Hsu, "Face liveness detection using a flash against 2D spoofing attack," *IEEE Trans. Inf. Forensics Secur.* **13** (2), 521–534 (2018).
28. Olarik Surinta and Thananchai Khamket, "Gender recognition from facial images using local gradient feature descriptors," in *2019 14th International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP)* (IEEE, 2019).
29. F. A. Pujol, M. J. Pujol, C. Rizo-Maestre, and M. Pujol, "Entropy-based face recognition and spoof detection for security applications," *Sustainability* **12** (1), 85 (2019).
30. Dhiman Karmakar and C. A. Murthy, "Face recognition using face-autocropping and facial feature points extraction," in *2nd International Conference on Perception and Machine Intelligence (PerMin), CDAC, Kolkata, India* (2015), pp. 116–122.

Dhiman Karmakar is presently acting as an Assistant Professor cum Head of the department, Computer Science, Surendranath college. He has also been a member of Post Graduate Board of studies, University of Calcutta. Formerly, he served MAKAUT, WB as an Assistant Professor cum Head of the department, Information Technology, for approximately a period of one year on lien. Erst he was associated in a research project on "color face recognition" in Machine Intelligence unit of Indian Statistical Institute, Kolkata. He did his graduation in Computer Sc. (Hons.) from University of Calcutta. He received Master of Computer Application in 2005 and M.Tech. In 2007 (in Computer Sc. and Eng.) degrees, both from University of Calcutta. He has qualified UGC NET in Computer Sc. & Applications and consequently ranked in WBCSC. He was also selected for CSIR Senior Research Fellowship in Computer Engineering Sciences. He had supervised projects of Msc (CS), Bsc Hons (CS), and B.Tech (CSE & IT) students. His research interests include computer vision, multi-modal biometrics, image steganography, speech/ speaker recognition, etc. and he has so far published five number of papers in reputed journals and international conferences.

Puja Mukherjee is presently a College Teacher in the Computer Science department of Surendranath college, Kolkata. She did her M.Tech in Computer Science and Engineering from Netaji Subhash Engineering College, WBUT in 2011. She did her B.E in Information Technology from University Institute of Technology, Burdwan University in 2009. She was engaged as an Assistant professor in the department of Computer Science and Engineering, Dream Institute of Technology, Kolkata from 2011. For the last couple of years she was involved in her doctoral work entitled "Adaptive Biometric System" in the Computer Science and Engineering department of University of Calcutta. Her research domain include image processing, pattern recognition, computer vision, etc.

Dr. Madhura Datta is presently an Assistant Professor in UGC-Human Resource Development Center of the University of Calcutta. She did her Ph.D (Tech.) degree on "Some Studies on Face Detection and Recognition" from University of Calcutta. She obtained M.Sc degree in Computer and Information Science from the University of Calcutta in 2002 and M.Tech degree in Computer Science and Engineering from the same University in 2004. She wrote a book chapter in the area of Biometric Recognition Systems. She has coauthored a book on Face Detection and Recognition published under CRC press. Her research interests include pattern recognition, 3D image analysis, machine learning, etc. Dr. Datta has more than fifteen research publications in reputed international journals and conferences.