Expert Systems With Applications 187 (2022) 115883

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

PATRON: Exploring respiratory signal derived from non-contact face videos for face anti-spoofing

Lokendra Birla [*], Puneet Gupta

Indian Institute of Technology Indore, Indore, India

ARTICLE INFO

Keywords:
Face anti-spoofing
r-PPG
Face authentication systems
Respiratory signal
3D face mask

ABSTRACT

Face authentication provides non-contact, user-friendly, covert, and low-cost acquisition. Despite this, face authentication is avoided in safety-critical applications because an adversary can easily spoof it. Several methodologies have been explored in the literature, but all of them, including remote Photoplethysmography (rPPG), are insufficient to detect the 3D face mask. The 3D face mask attack is considered the most potent attack, and it cannot be correctly detected even after consolidating different methodologies. It motivates us to explore a different methodology for face anti-spoofing based on respiration rate because it provides complementary information with the existing methodologies. To achieve the best possible performance, our novel method, $PATRON$ that is res**P**iration b**A**sed fea**T**u**R**es f**O**r 3D face mask a**N**ti-spoofing is based on: i) different characteristics as that of rPPG methods; ii) appropriate selection of facial regions; iii) relevant feature selection; and iv) compact feature representation. Our extensive experimental results on a publicly available 3D face mask anti-spoofing dataset reveal that our proposed method $PATRON$ performs similar to the several state-of-the-art methods, and respiration rate can be utilized for face anti-spoofing. Furthermore, it provides guidelines about proper facial region selection and feature extraction, which enables the respiratory signal for anti-spoofing.

1. Introduction

In the modern digital age, the face authentication system has been widely used by humans to perform their day-to-day activities, pertaining to smartphone face unlocking, robotics, human–computer interaction, healthcare, and surveillance. The factor behind this success is that face can be acquired in a non-contact and user-friendly manner from a low-cost acquisition sensor. Despite these advantages, face authentication is avoided in safety-critical applications involving economic impacts because the facial images can be easily obtained from social media and utilized by an adversary to spoof (or fool) the face authentication system. It motivates us to design a face anti-spoofing method, which will accurately distinguish between genuine and spoofed faces.

In face spoofing, an adversary aims to impersonate a genuine person's identity by fooling the face recognition system. One of the simplest ways to perform face spoofing is using a print attack (or photo attack). In this attack, the adversary prints the photo of a genuine person's face and provides this photo to the acquisition device for fooling the system. Since the photo contains a static face devoid of any facial movements, this attack can be detected by motion-based face anti-spoofing methods, which analyze the facial movements (Feng et al., 2016). The motion-based anti-spoofing methods are ineffective for a video replay attack or an eye-cut photo attack. In a video replay attack, the adversary gets a face video of a genuine person from social media and plays it in front of the face recognition system. In an eye-cut photo attack, the adversary cuts the portion of the photo's eyes, hides his/her face behind the photo, and blinks his/her eyes during face acquisition. When the adversary uses a low-quality image of the genuine user, these attacks can be detected using the appearance-based face anti-spoofing methods (Maatta, Hadid, & Pietikinen, 2011). The appearance-based methods analyze the texture and image quality for face anti-spoofing (Liu et al., 2019).

The print attack, replay attack, and eye-cut photo attack rely on utilizing the 2D spoofed faces. Thus, these attacks are devoid of depth information. This intuition is leveraged in the literature by consolidating depth information with other anti-spoofing methods to improve the efficacy (Atoum, Liu, Jourabloo, & Liu, 2017). Unfortunately, it is possible to make the depth information redundant in face anti-spoofing by warped photo attack where a bent printed photo is shown to the acquisition device for fooling the system (Wang, Nian, Li, Meng, & Wang, 2017). Apart from these attacks, an adversary can perform face spoofing by a 3D mask attack where a genuine user's 3D face mask is placed on the adversary's face. The 3D face mask is flexible for

[*] Corresponding author.
E-mail addresses: phd1901201001@iiti.ac.in (L. Birla), puneet@iiti.ac.in (P. Gupta).

https://doi.org/10.1016/j.eswa.2021.115883
Received 10 January 2021; Received in revised form 6 July 2021; Accepted 5 September 2021
Available online 20 September 2021

simulating facial gestures and simultaneously contains genuine human face-like texture and depth information. Thus, the traditional motion, appearance, or depth-based face anti-spoofing methods are insufficient for detecting the 3D face mask attack, and this attack is considered the most potent attack (Li, Komulainen, Zhao, Yuen, & Pietikainen, 2016).

The 3D face mask attack can be detected by remote Photoplethysmography (rPPG) technique (Li et al., 2016; Liu, Lan and Yuen, 2018; Liu, Lan, & Yuen, 2020; Liu, Yuen, Zhang, & Zhao, 2016; Nowara, Sabharwal, & Veeraraghavan, 2017). This technique estimates the blood flow variations inside the facial arteries of the human face. The estimation can be achieved by extracting the subtle facial color and motion variations from the non-contact face videos acquired from a camera (Guo et al., 2020). These variations are latent to the human eye but fortunately captured by the camera. Since this technique estimates the blood flow variations from the human face, it provides random noise when a 3D face mask covers the face. Face anti-spoofing using the rPPG technique leverage this observation. One major factor that restricts the efficacy of rPPG based methods is that the extracted heart rate information can be easily affected by facial expressions, illumination, and camera parameters (Gupta, Bhowmik, & Pal, 2018b).

Blood flow variations contain relevant information about vital human parameters. However, the existing rPPG based anti-spoofing methods utilize only the heart rate information and neglect the remaining information related to respiration rate and several other human physiological parameters. Unlike heart rate, which affects all the facial regions, the information of respiration rate is prominent in small facial regions, and thus, inferring the respiration rate is a difficult task compared to the heart rate (Cattani et al., 2017). Despite being a more challenging task, the respiration rate provides complementary information with heart rate. Thus, it can be used along with existing rPPG based methods to improve the efficacy. Motivated by this observation, we propose the novel face anti-spoofing method, $PATRON$ that is resPiration bAsed feaTuRes fOr 3D face mask aNti-spoofing. It explores the feasibility of respiratory signals for 3D face mask detection. The main contributions of this paper are:

1. To the best of our knowledge, this is the first research where the respiratory signal is utilized for 3D face mask anti-spoofing. Our study shows that the respiratory signal provides valuable information required for anti-spoofing.
2. The respiration and heart rate can be estimated from non-contact face videos using a similar phenomenon. Thus, it can be apparent that face anti-spoofing using respiratory signal can be performed using the methodology prevalent for the rPPG based anti-spoofing methods. However, we demonstrate that the anti-spoofing based on heart rate and respiratory signal utilizes contradictory principles.
3. The performance of respiratory signal-based anti-spoofing is dependent on the region of interest (ROI) selection and temporal signal extraction. These parameters are thoroughly analyzed, and the most appropriate ROI and temporal signal extraction methods are provided to achieve the best possible 3D face mask anti-spoofing.
4. Our novel face anti-spoofing method, $PATRON$ is capable of providing a compact feature representation, which is essential for learning the model parameters using limited training examples (Chingovska, Anjos, & Marcel, 2012).

The organization of this paper is as follows. A brief description of the existing works that are required to augment the understanding of the proposed method is provided in the next section. Subsequently, the proposed method, $PATRON$ is presented in Section 3. It is followed by understanding the experimental results. Conclusions are provided in the last section.

2. Related work

This section provides a brief description and limitations of well-known anti-spoofing methods, namely, motion-based, appearance-based, rPPG based, and hybrid face anti-spoofing methods. Subsequently, it outlines existing methods for respiration rate estimation using non-contact face video because the proposed method aims to provide a new direction of utilizing the respiration rate for face anti-spoofing.

2.1. Existing anti-spoofing methods

Face recognition systems can be spoofed using print photo attacks, video replay attacks, eye-cut photo attacks, bent photo attacks, and 3D face mask attacks. Several face anti-spoofing methods have been proposed in the literature that aims to detect these spoofed faces. Unfortunately, the efficacy of these anti-spoofing methods is limited.

2.1.1. Motion based anti-spoofing methods

The simplest possible face spoofing is to utilize the printed photograph of a genuine face. The photograph contains a static genuine face. Hence, the movements between the different facial regions are similar. In contrast, the genuine face contains dissimilar movements between the different facial regions due to facial expressions and inevitable eye-blinking. Thus, the motion-based face anti-spoofing methods analyze the movements between the different facial regions for detecting photograph-based attacks. Some such well known methods are based on analyzing the eye blinking (Kollreider, Fronthaler, & Bigun, 2008), mouth movement (Kollreider, Fronthaler, Faraj, & Bigun, 2007), facial expression (de Freitas Pereira et al., 2014) and head movement (Feng et al., 2016). Usually, the optical flow is employed for estimating the movements due to eye blinking and facial expressions (Kollreider et al., 2008). The movements between the different facial regions can also be estimated using the correlation between the background and facial region (Yan, Zhang, Lei, Yi, & Li, 2012). The motion-based anti-spoofing methods can be easily spoofed using the video replay attack, eye-cut photo attack, and 3D mask attack.

2.1.2. Appearance based anti-spoofing methods

An adversary can use the printed genuine face to fool the face recognition system in several spoofing attacks. The material where a genuine face is printed contains different texture and reflection characteristics compared to that of a genuine face. This observation is leveraged in appearance-based face anti-spoofing methods for detecting print photo and warped photo attacks. These methods analyze the facial texture and image quality for distinguishing between the spoofed and genuine face (Li, Wang, Tan, & Jain, 2004). Some such well known texture-based methods are multiscale LBP (Boulkenafet, Komulainen, & Hadid, 2016b), SIFT (Patel, Han, Jain, & Ott, 2015), and SURF (Boulkenafet, Komulainen, & Hadid, 2016a) and color-texture (Jourabloo, Liu, & Liu, 2018). Likewise, face anti-spoofing can be performed by analyzing the image quality defects between spoofed and genuine faces. To this end, reflection pattern (Wen, Han, & Jain, 2015) and moire patterns (Patel et al., 2015) can be utilized. Kindly note that the video replay and eye-cut photo attacks are undetected using appearance-based methods when the adversary contains a high-quality image of the genuine user (Maatta et al., 2011).

2.1.3. rPPG based anti-spoofing methods

Traditionally, the cardiovascular pulse is estimated using ECG, EEG, pulse-oximeter, or wearables. They require skin contact, which restricts their applicability in several real-world applications (Gupta, Bhowmick, & Pal, 2017a, 2018a; Hamedani, Bahmani, & Mohammadian, 2016; Hirsch, Jensen, Poulsen, & Puthusserypady, 2020). It is possible to extract the cardiovascular pulse using rPPG methods where the non-contact face video acquired from the camera is utilized for

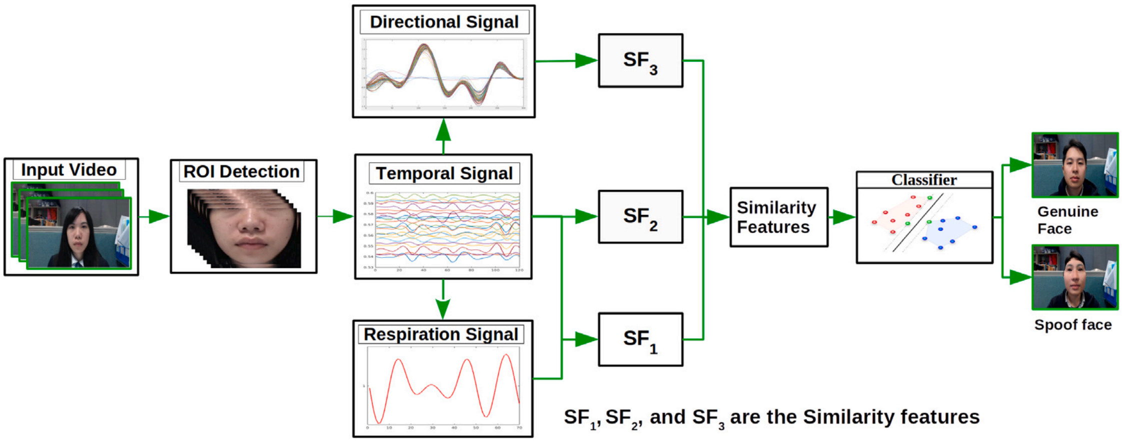


Fig. 1. Flow Diagram of our proposed 3D face mask anti-spoofing method, $PATRON$.

the extraction. The rPPG methods are based on the phenomenon that the heartbeat introduces blood flow variations (or color variations) and head movements (or motion variations), which are subtle enough to be latent to the human eye but can be observed from the camera videos (Gupta, Bhowmick, & Pal, 2017b). These are categorized as either the Eulerian approach, which observes the color variation in a fixed ROI amongst subsequent frames (Gupta, Bhowmick, & Pal, 2020) or the Lagrangian approach, which observes the movements of discriminatory features in the subsequent frames (Balakrishnan, Durand, & Guttag, 2013).

Pulse signal can be accurately extracted from the non-contact face video of the genuine face, and this signal has a high signal-to-noise ratio (SNR) (Gupta et al., 2018a). In contrast, pulse signal cannot be detected when the face is covered with 3D face masks because the mask obstructs the acquisition of facial color or blood flow variations in the camera (Li et al., 2016). In such a case, the signal extracted from the non-contact face video contains noise having a low SNR value rather than the pulse information. This observation is utilized in Li et al. (2016) and Nowara et al. (2017) for performing the face anti-spoofing based on SNR. The 3D face mask anti-spoofing can also be performed by analyzing the correlation between the pulse signals extracted from different facial regions (Liu, Lan et al., 2018; Liu et al., 2020, 2016). Based on the observations, this correlation is high for genuine faces because the pulse signals extracted from different facial regions correspond to cardiovascular pulse signals and some noise. In contrast, this correlation is low for spoofed faces where noise signals are usually extracted rather than pulse signals.

The noise characteristics and heart rate in rPPG methods share a similar frequency range (Wang, den Brinker, Stuijk, & de Haan, 2016) and thus, the extracted pulse signal can be easily affected by facial expressions, illumination, and camera parameters (Kotwal, Bhattacharjee, & Marcel, 2019). Due to these factors, rPPG methods have limited applicability for 3D mask face anti-spoofing because noisy pulse signals will eventually result in erroneous face anti-spoofing.

2.1.4. Hybrid anti-spoofing methods

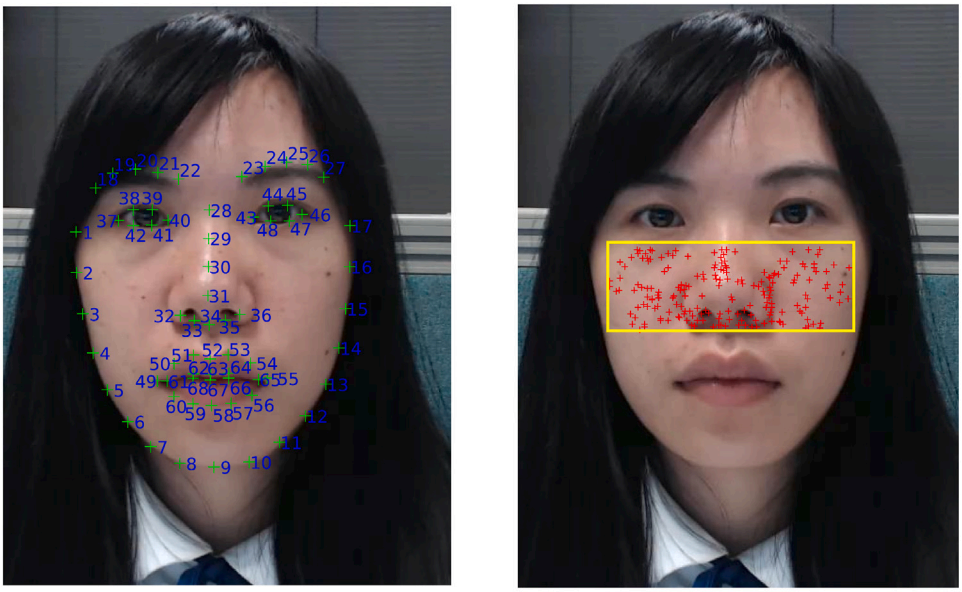
Several anti-spoofing methods exist in the literature, but each method has limited applicability for detecting the 3D face mask. The existing methods are consolidated to improve the face anti-spoofing. Such methods are referred to as hybrid methods. In most of these methods, the depth information is consolidated with other anti-spoofing methods to detect those attacks that utilize 2D spoofed faces. Some such attacks are the print attack, replay attack, and eye-cut photo attack. However, they are ineffective against 3D face mask attacks because the acquired spoofed face contains depth information. Some hybrid methods where the depth information is consolidated with the appearance or rPPG methods are Atoum et al. (2017), Liu, Jourabloo and Liu (2018) and Wang et al. (2017). In Atoum et al. (2017) and Wang et al. (2017), the depth information and texture features are consolidated for anti-spoofing. In contrast, face anti-spoofing is performed by consolidating the depth information and pulse signal information in Liu, Jourabloo et al. (2018).

2.2. Respiration rate using face videos

Like the cardiovascular pulse signal, the respiratory signal can be estimated from non-contact face videos by observing the subtle blood flow variations (or color variations) and head movements (or motion variations). The respiratory signal is estimated in Luguern et al. (2020) and Sanyal and Nundy (2018) by analyzing the subtle blood flow or color variations. In Chen, Zhu, Zhang, Wu, and Wang (2019), the subtle head movements are estimated by tracking the discriminatory facial point. Such methods can provide incorrect respiration rate when: (i) the input face video is affected by facial motions and expressions; (ii) facial region is not properly selected, and (iii) there are variations in extrinsic parameters like illumination and camera parameters. To the best of our knowledge, the respiratory signal is not utilized for face anti-spoofing, and we are the first to propose a 3D face mask anti-spoofing method based on the respiration signal obtained from non-contact face videos.

3. Proposed method

The proposed 3D face anti-spoofing method, $PATRON$, is presented in this section. Its flow-graph is shown in Fig. 1. It aims to determine whether the person in the video is a genuine person or a spoofed person. To this end, the respiratory signal of the person is estimated. Initially, the facial region containing a prominent respiratory signal is extracted from the face video, and the extracted region is referred to as ROI. Subsequently, the temporal deformations in the ROI for all the subsequent video frames are estimated. The resultant signal is known as a temporal signal. The post-processing techniques are applied to the extracted temporal signals, which mitigate the noise and pulse signal while preserving the respiratory signal. Subsequently, the respiratory signal is extracted from the resultant signals. The relevant features are extracted from the temporal signals and respiratory signal. These features are eventually provided to the Support Vector Machine (SVM) classifier for face anti-spoofing. Since limited training data is available in the realm of 3D face anti-spoofing, the classifier is prone to overfitting and can provide erroneous results (Rehman, Po, & Liu, 2018, 2020). To mitigate this issue, we define our compact features and simultaneously preserve most of the relevant information.



(a) Landmark Detection.

(b) Extracted ROI and feature points.

Fig. 2. Example of ROI and feature extraction.

3.1. ROI detection

Our method performs facial analysis for face anti-spoofing. Thus, it first detects the face present in the input video and removes the remaining background or non-facial area for further analysis. We employ CLNF OpenFace 2.2.0 (Zadeh, Chong Lim, Baltrusaitis, & Morency, 2017) for the face detection. It also provides the facial landmark points, which outline the facial boundary, eye-brow regions, eye regions, mouth regions, and nose region. An example of such facial landmarks can be visualized from Fig. 2(a). We use the landmark points for extracting the ROI, which, in our case, denotes the facial region containing a significant respiratory signal. Usually, the nose region is used amongst the facial regions to extract the respiratory signal because it is least affected by non-rigid facial deformation (Gupta et al., 2018b) and contains the prominent respiratory signal (Bennett, Goubran, & Knoefel, 2017). In contrast, the eyes, forehead, mouth, and facial boundaries are avoided because these regions are easily affected by facial expressions and do not contain significant information regarding the respiratory signal. We leverage these observations by defining our ROI such that it encloses the nose and nearby facial regions while avoiding eyes, forehead, mouth, and facial boundaries. For this purpose, we obtain the ROI by computing the minimum enclosing area rectangle (Boltz, Debreuve, & Barlaud, 2009) of the following landmark points: (i) 1, 2, and 3 of the left cheek; (ii) 33, and 35 of the nose; (iii) 15, 16, and 17 of the right cheek; and, (iv) 29 of the nose. There is subtle motion in the facial boundary pixels due to facial expressions. It can deteriorate the extraction of the temporal signal; thus, the boundary pixels are removed by performing the morphological erosion (Gupta & Gupta, 2015). An example of ROI extraction can be visualized from Fig. 2(b), whereas the ROI is shown in the yellow rectangle. We further explore the efficacies of different ROI selections, and the experimental results are discussed in Section 4. The results indicate that our ROI is most suitable for 3D mask face anti-spoofing.

3.2. Temporal signal extraction

In this subsection, the temporal signals are estimated, which represent the temporal deformations in the extracted ROI for all the video frames. These deformations can be estimated using either Eulerian or Lagrangian approaches. Amongst these approaches, the Lagrangian approach provides better temporal signals, and thus, we extract the temporal signals using the Lagrangian approach. For this purpose, we first detect the discriminatory points in the ROI using the minimum eigenvalue algorithm (Shi et al., 1994). For illustration, an example of these discriminatory facial points is shown using red dots in Fig. 2(b). These points are tracked in subsequent frames using the Kanade–Lucas–Tomasi (KLT) algorithm (Kalal, Mikolajczyk, & Matas, 2010). Thus, two temporal signals are extracted from each discriminatory point which provides the deformations in x and y directions. Mathematically, assume that the tracked movements of Lth points is given by $\left[\left(X_L^1, Y_L^1\right), \left(X_L^2, Y_L^2\right), \ldots, \left(X_L^N, Y_L^N\right)\right]$ where N is number of frames in a video and $\left(X_m^i, Y_m^i\right)$ represents the location of mth point in ith frame. Then, the corresponding temporal signals in the vertical and horizontal direction represented by T_L^x and T_L^y, respectively, are given by:

$$T_L^x = \left(X_L^1, X_L^2, \ldots, X_L^N\right) \text{ and } T_L^y = \left(Y_L^1, Y_L^2, \ldots, Y_L^N\right) \tag{1}$$

The extracted temporal signals contain the respiratory signal along with the pulse signal and noise. The noise is induced by facial expressions, illumination variations, and camera sensor noise. We apply the post-processing technique to the extracted temporal signals for mitigating the noise and pulse signal while preserving the respiratory signal. We apply a bandpass filter as a post-processing technique to each temporal signal. Its cutoff frequencies are set from 0.083 (5/60) to 0.5 (30/60) Hz. Since the respiration rate lies between 5–30 Breaths per minute (BPM) (Chen et al., 2019), our bandpass filter preserves the respiratory signal in the resultant temporal signals while removing the noise and pulse signal lying outside the range of 5–30 BPM.

3.3. Respiratory signal extraction

Each resultant temporal signal obtained after the post-processing mainly contains the respiratory signal along with the noise. Usually, the blind source separation (BSS) is employed in such scenarios where the source signal (which is the respiratory signal in our case) needs to be estimated from the output signals (which are temporal signals in our case) (Gupta et al., 2018b). Furthermore, the human vital parameters should be periodic (Balakrishnan et al., 2013). The BSS and high periodicity are simultaneously used in Gupta et al. (2018b) to propose a Multi-kurtosis optimization. We utilize this Multi-kurtosis optimization (Gupta et al., 2018b) for extracting the respiratory signal from the resultant temporal signals.

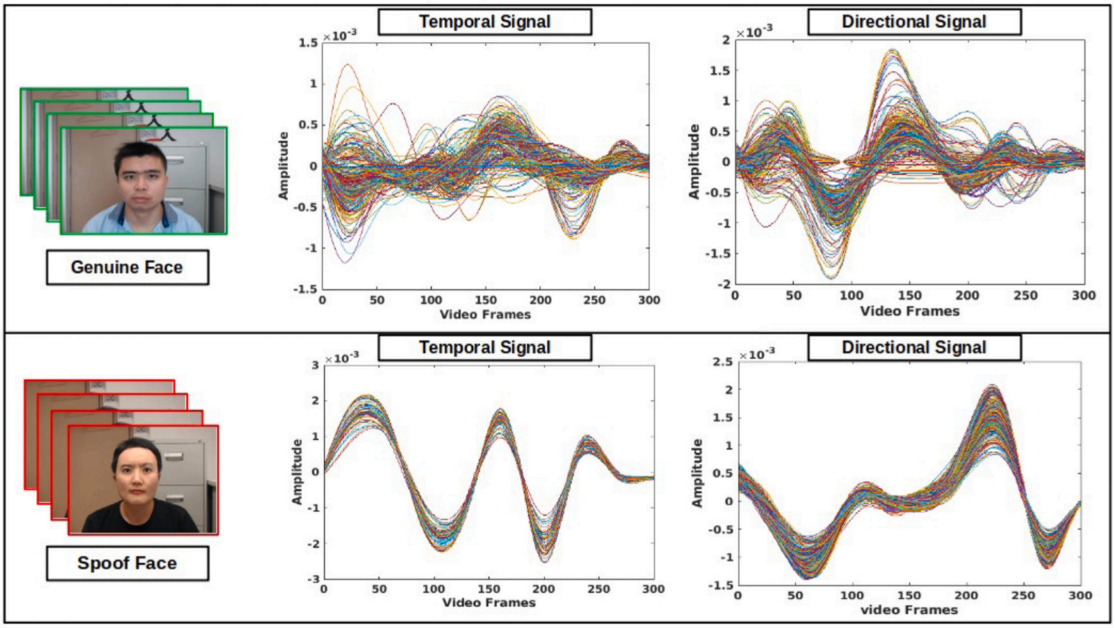


Fig. 3. Behavior of temporal and directional signals for Genuine and spoofed cases. It depicts the temporal and directional signals of two genuine faces in the upper rectangle and two spoofed faces in the lower rectangle. It depicts that the temporal and directional signals are less correlated for genuine faces in the above rectangle box. In contrast, the temporal and directional signals in spoofed faces exhibit higher correlation, as shown in the lower rectangle.

3.4. Feature extraction

One can train a classifier for face anti-spoofing by providing the respiratory signal as input. This classifier will provide low performance because (i) respiratory signals for different subjects have variable amplitudes, and (ii) the number of parameters required by such a classifier is large as the respiratory signal is large, but the training data is limited. To mitigate the problem of different respiratory signals' amplitudes for different people, we define the three similarity features using the correlation between temporal signals and respiratory signals. Subsequently, these similarity features are compactly represented to handle the problem of limited training data.

The intuition behind using the similarity features is that a 3D face mask is a rigid object placed over the face. Hence, the temporal movements of the different points on the mask are similar. In other words, there is a high correlation between the temporal signals and respiratory signal for 3D face mask spoofed faces. In contrast, a genuine face contains elastic skin that can stretch and contract; thus, temporal movements of different facial points are different. Hence, a low correlation between temporal signals and the respiratory signal can be expected for genuine faces as compared to spoofed faces. It is important to understand that this intuition contradicts with the behavior prevalent in rPPG based face anti-spoofing methods (Liu, Lan et al., 2018; Liu et al., 2020, 2016). In rPPG methods, the correlation is high and low for genuine and spoofed faces, respectively, because (i) the temporal signals extracted from different facial regions of a genuine face mainly contain the pulse signal along with some noise, and (ii) this correlation is low for spoofed faces where random noise signals are usually extracted from different facial regions while pulse signal is attenuated due to mask. The contradictory behavior in rPPG methods and respiratory signal-based face anti-spoofing can be attributed to the different noise characteristics and frequency ranges.

3.4.1. First similarity feature (SF_1)

Our first similarity feature is defined using the normalized correlation between temporal signals and respiratory signal. It is observed in Balakrishnan et al. (2013) that in the Lagrangian approach, the vertical facial movements mainly contain the pulse information while most of the respiratory information is present in the horizontal facial movements. Thus, we only utilize the temporal signals belonging to the horizontal direction. Mathematically, if there are p discriminatory points, then there are p temporal signals belonging to the horizontal directions given by $[T_1^x, T_2^x, \dots, T_p^x]$ where T_i^x (refer Eq. (1)) provides the horizontal temporal signal corresponding to the ith facial discriminatory point. Using these horizontal temporal signals, total p normalized correlations given by $[m^1, m^2, \dots, m^p]$ are computed where m^i denotes the normalized correlation between the respiratory signal, R and the ith temporal signal, T_i^x. That is, m^i is given by:

$$m^i = \frac{(R)^T (T_i^x)}{\|(R)^T\|_2 \left\|(T_i^x)\right\|_2} \tag{2}$$

where $\|\cdot\|_2$ denotes the l_2 norm.

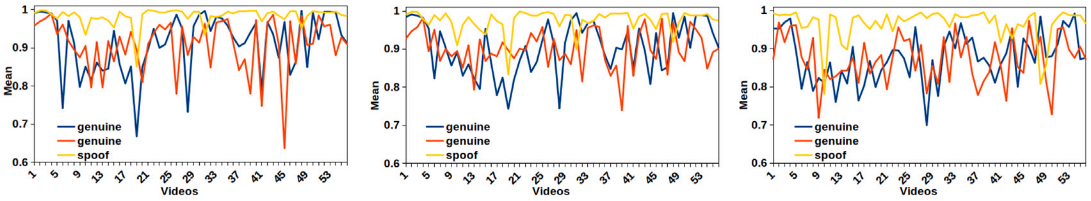
It can be observed that the number of normalized correlations depends on the number of distinctive points selected in the ROI, which are variable and usually large in number. Unfortunately, there are limited training samples available in our case. Hence, our trained model will suffer from the problem of overfitting when these large dimensional normalized correlations are used as features. We mitigate this issue by compactly representing these correlations using their mean and use it as our first feature. Mathematically, our first feature SF_1 is given by:

$$SF_1 = \frac{\sum_{i=1}^{p} m^i}{p} \tag{3}$$

where p is the number of temporal signals in a horizontal direction. An example depicting SF_1 feature is presented in Fig. 4a. The yellow line denotes the SF_1 corresponding to spoofed face videos, whereas the other two plots in the red and blue line denote the SF_1 corresponding to genuine face videos. It can be observed that the correlation is usually higher for spoofed videos as compared to genuine videos.

3.4.2. Second similarity feature (SF_2)

Our second similarity feature is defined using the normalized correlation between temporal signals in the horizontal direction as these signals provide more information about respiratory signal than vertical direction (Balakrishnan et al., 2013). The temporal signals in the horizontal direction are better correlated for the spoofed face than



(a) SF_1 (b) SF_2 (c) SF_3

Fig. 4. Behavior of similarity features for Genuine and Spoof faces. The x-axis denotes the label or number of a video. In contrast, the y-axis denotes the magnitude of the similarity feature.

the genuine face. An example depicting this phenomenon is shown in Fig. 3. The figure shows that the temporal signal in the spoofed faces exhibits a higher correlation than the genuine faces. Mathematically, the normalized correlation between temporal signals is computed using Eq. (2). Furthermore, if there p horizontal temporal signals, then there are ${}^{p}C_2$ normalized correlations which are large in number. Thus, we compactly represent the correlations using their mean and use it as our second feature. Mathematically, our second feature SF_2 is given by:

$$SF_2 = \frac{\sum_{i=1}^{p}\sum_{j=i+1}^{p} r_x^{ij}}{{}^{p}C_2} \tag{4}$$

where p is the number of temporal signals in horizontal direction and r_x^{ij} denote the normalized correlation between ith and jth horizontal temporal signals which are given by T_i^x and T_j^x, respectively. An example depicting the SF_2 feature is presented in Fig. 4b in which the yellow line represents SF_2 of spoof faces, whereas the other two plots show the low correlation of genuine faces.

3.4.3. Third similarity feature (SF_3)

Intuitively, all the facial feature points move in a similar direction for spoof faces, but facial feature points move in different–different directions due to skin elasticity in the case of genuine faces. Thus, the directional movements of distinctive points in the ROI provide useful information about anti-spoofing. This observation motivates us to incorporate the directional movements of distinctive points (Gupta & Gupta, 2016). This behavior is illustrated in Fig. 4a. The figure shows that the directional signal in the spoofed faces exhibits a higher correlation than the genuine faces in most cases. Mathematically, directional signal for the ith point, D_i is given by:

$$D_i = \left(tan^{-1}\left(\frac{T_i^y(1)}{T_i^x(1)}\right), tan^{-1}\left(\frac{T_i^y(2)}{T_i^x(2)}\right), \dots, tan^{-1}\left(\frac{T_i^y(q)}{T_i^x(q)}\right)\right) \tag{5}$$

where q denotes the length of temporal signals while T_i^y and T_i^x are temporal signals in the vertical and horizontal axis, as mention in Eq. (1). Our third similarity feature is defined using the normalized correlation between the directional signal. It can be observed that if there are p directional signals, then there are ${}^{p}C_2$ normalized correlations which are large in number. Thus, just like SF_1 and SF_2, we compactly represent the correlations using their mean and use it as our third feature. Mathematically, the third similarity feature, SF_3 is given by:

$$SF_3 = \frac{\sum_{i=1}^{p}\sum_{j=i+1}^{p} s_x^{ij}}{{}^{p}C_2} \tag{6}$$

where p is total number of temporal signals or directional signals in the input video while s_x^{ij} denotes the normalized correlation between ith and jth directional signals which are given by D_i and D_j, respectively. An example depicting SF_3 feature is presented in Fig. 4c. It demonstrates that the spoofed faces usually have higher values (shown in yellow color) as compared to genuine faces (shown in blue and red color).

3.5. Classification

Our similarity features SF_1, SF_2 and SF_3 obtained from Eqs. (3), (4), and (6), respectively, these similarity features are concatenated into one vector and provided to our classifier for performing the 3D mask face anti-spoofing. We use SVM based classifier with RBF kernel for this purpose.

4. Experiments

4.1. Dataset

We evaluated our proposed method's performance, $PATRON$, using a publicly available 3D mask attack dataset HKBU-MARsV1+ (Liu, Lan et al., 2018). It comprises 180 videos, out of which 120 videos contain genuine face videos and the remaining 60 videos contain 3D mask spoofed faces. The videos are acquired from 12 subjects, where each subject provides 15 videos, out of which 10 videos are genuine face videos, and the remaining 5 videos are the 3D mask face videos. The face resolution of all videos is 200 × 200 pixels. The spoofed videos are acquired by utilizing two types of masks, which are TMF masks and high-quality masks from REAL-f.

4.2. Implementation details

Our proposed method is implemented in MATLAB 2019a and tested on the Intel i5-5200U CPU@2.2 GHz processor. Furthermore, our classifier is designed using SVM with the RBF kernel using the LIBSVM library. We also adopted the Liu, Lan et al. (2018) methodology and computed the true positive rate (TPR) and false-positive rate (FPR) at different threshold values. TPR and FPR at different thresholds are used to create the Receiver Operating Characteristics (ROC) curve. The area under the ROC curve is computed and used as a performance metric, AUC in the experiments. Likewise, Equal Error Rate (EER) is used as another performance metric in the experiments. The EER is given by the FPR at that threshold where FPR and (1-TPR) are equal.

4.3. Testing protocol

The proposed method $PATRON$ is tested with the state-of-the-art methods under the same settings for comparison. We perform the experiments on the HKBU-MARsV1+ dataset using the leave one out cross-validation (LOOCV) methodology (Liu, Lan et al., 2018). This dataset consists of 11 subjects, and data of subject number 8 was not provided due to privacy issues; hence we reimplemented the state-of-the-art methods for comparative analysis of our proposed method.

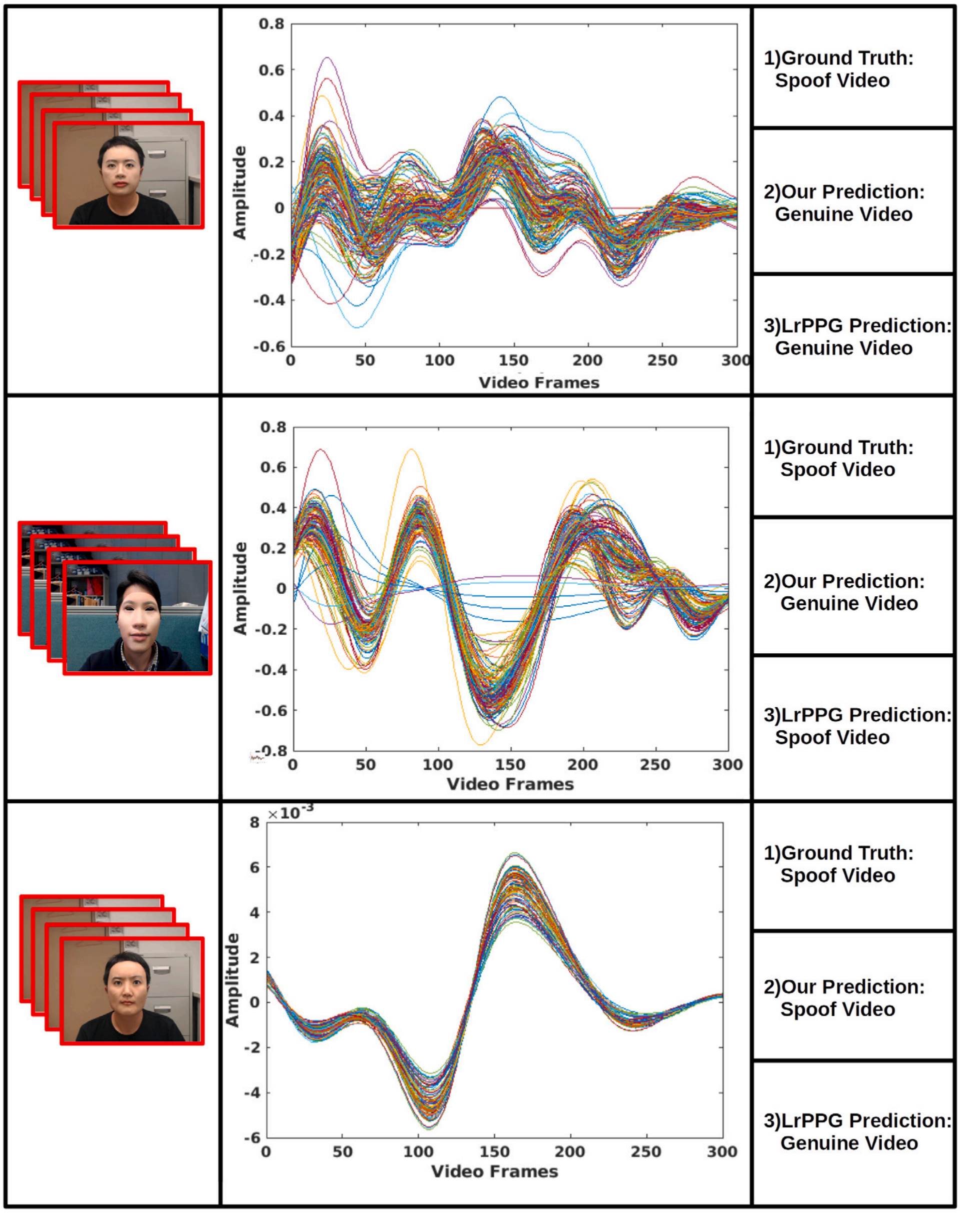


Fig. 5. Qualitative results of the proposed method. In this figure, we demonstrate three examples of spoofed faces using the temporal signal. Each row represents a sample with corresponding temporal signals and predictions. The figure depicts the wrong and correct prediction of the samples by our proposed method and another state-of-the-art method, LrPPG. The first row depicts a case where both our proposed method and LrPPG provide the wrong prediction because temporal signals are poorly correlated due to face movement. The second row depicts a case where temporal signals are not correlated, and our method mispredicts it, but LrPPG predicts it correctly. The last row depicts the case where temporal signals are highly correlated, and our method predicts it correctly, whereas LrPPG predicts it wrong.

Table 1
Comparative performance evaluation of our proposed method, $PATRON$.

	AUC	EER
OFF (Bao, Li, Li, & Jiang, 2009)	46.5	52.3
MS-LBP (Erdogmus & Marcel, 2014)	85.8	22.5
CTA (Boulkenafet, Komulainen, & Hadid, 2015)	82.3	23.0
sCNN-LE (Qu, Dong, & Niu, 2019)	86.2	16.5
Shao, Lan, and Yuen (2018)	86.9	17.8
MS-LTSS (Lin, Li, Yu, & Zhao, 2019)	85.7	18.4
GrPPG (Li et al., 2016)	85.2	17.3
LrPPG (Liu et al., 2016)	87.3	15.2
CFrPPG (Liu, Lan et al., 2018)	88.2	14.8
$PATRON$	**87.8**	**14.7**

4.4. Comparative results

In this subsection, our proposed method's performance, $PATRON$, is compared with several state-of-the-art methods. The comparative results are provided in Table 1. Even though several rPPG methods exist in the literature that utilizes cardiovascular pulse for anti-spoofing, according to the best of our knowledge, we are the first one to explore the utilization of respiratory signals for 3D face mask detection. Fortunately, our choice is properly justified by the results depicted in Table 1. The table indicates that our respiratory signal-based method, $PATRON$ provides better results than well-known texture-based methods (Bao et al., 2009; Boulkenafet et al., 2015; Erdogmus & Marcel, 2014; Qu et al., 2019; Shao et al., 2018). The possible reasons for this phenomenon are that the texture-based methods suffer from the overfitting problem due to data-driven property (Liu et al., 2016) and 3D face mask contains similar texture as that of the human face.

Table 2
Experimental Analysis of our proposed method $PATRON$.

	ROI	FR[a]	Similarity features			Lagrangian approach		Eulerian approach[b]	
			SF_1	SF_2	SF_3	AUC	EER	AUC	EER
$PATRON$	R_1	✗	✓	✓	✓	**87.8**	**14.7**	NA	NA
1	R_1	✗	✓	✗	✗	76.1	27.5	63.0	40.3
2	R_1	✗	✗	✓	✗	69.3	34.7	62.2	41.1
3	R_1	✗	✗	✗	✓	75.8	28.1	NA	NA
4	R_1	✗	✓	✓	✗	79.5	24.3	69.3	35.7
5	R_1	✗	✓	✗	✓	84.3	20.3	NA	NA
6	R_1	✗	✗	✓	✓	81.2	23.1	NA	NA
7	R_2	✗	✓	✗	✗	76.8	27.4	80.0	24.2
8	R_2	✗	✗	✓	✗	75.0	29.1	63.8	40.8
9	R_2	✗	✗	✗	✓	75.9	27.5	NA	NA
10	R_2	✗	✓	✓	✗	80.9	22.7	83.6	21.3
11	R_2	✗	✓	✗	✓	81.0	23.5	NA	NA
12	R_2	✗	✗	✓	✓	81.5	23.8	NA	NA
13	R_2	✗	✓	✓	✓	82.7	21.1	NA	NA
14	R_3	✗	✓	✗	✗	77.3	26.5	76.8	28.0
15	R_3	✗	✗	✓	✗	76.2	27.4	65.4	39.8
16	R_3	✗	✗	✗	✓	80.9	24.2	NA	NA
17	R_3	✗	✓	✓	✗	77.8	26.3	80.9	21.6
18	R_3	✗	✓	✗	✓	81.9	22.8	NA	NA
19	R_3	✗	✗	✓	✓	81.4	23.3	NA	NA
20	R_3	✗	✓	✓	✓	83.5	20.3	NA	NA
21	R_1	✓	✓	✗	✗	66.8	37.2	62.0	43.8
22	R_1	✓	✗	✓	✗	68.0	34.4	61.4	39.5
23	R_1	✓	✗	✗	✓	64.4	39.7	NA	NA
24	R_1	✓	✓	✓	✗	70.1	34.3	65.8	38.1
25	R_1	✓	✓	✗	✓	67.2	35.6	NA	NA
26	R_1	✓	✗	✓	✓	68.5	36.9	NA	NA
27	R_1	✓	✓	✓	✓	69.3	33.1	NA	NA
28	R_2	✓	✓	✗	✗	59.7	44.1	60.0	44.8
29	R_2	✓	✗	✓	✗	58.5	45.7	59.5	43.9
30	R_2	✓	✗	✗	✓	55.1	49.3	NA	NA
31	R_2	✓	✓	✓	✗	62.4	43.5	61.3	43.8
32	R_2	✓	✓	✗	✓	59.7	45.9	NA	NA
33	R_2	✓	✗	✓	✓	60.2	45.1	NA	NA
34	R_2	✓	✓	✓	✓	64.7	41.2	NA	NA
35	R_3	✓	✓	✗	✗	58.7	46.2	56.0	48.8
36	R_3	✓	✗	✓	✗	56.7	48.9	61.8	43.1
37	R_3	✓	✗	✗	✓	55.4	47.1	NA	NA
38	R_3	✓	✓	✓	✗	61.0	43.7	62.6	42.9
39	R_3	✓	✗	✓	✓	62.4	43.3	NA	NA
40	R_3	✓	✓	✗	✓	60.8	44.2	NA	NA
41	R_3	✓	✓	✓	✓	63.0	40.8	NA	NA

[a] FR denotes the face registration performed using OpenFace 2.2.0 (Zadeh et al., 2017).

[b] Eulerian approach cannot be used for SF_3 feature. Such cases are marked with NA: Not Applicable.

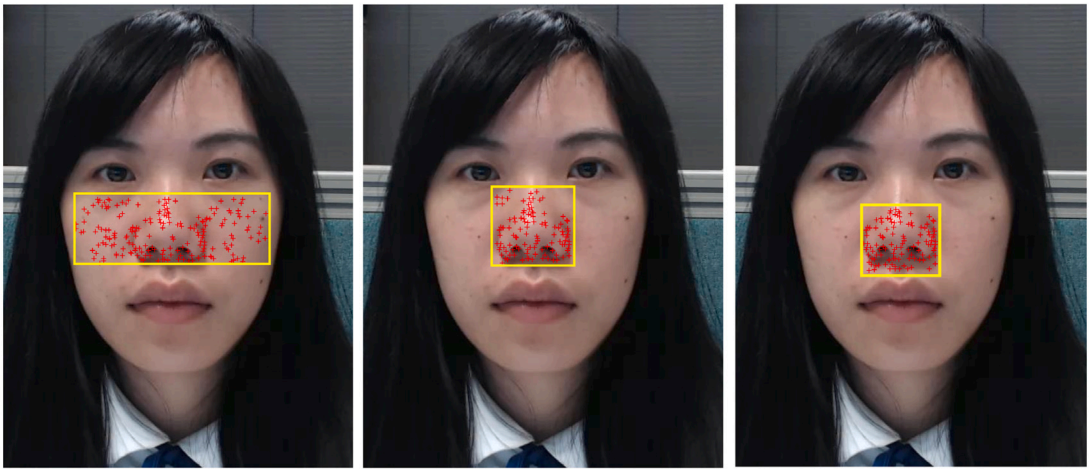
Furthermore, our respiratory signal based method, $PATRON$ performs similar to the several rPPG based methods, namely MS-LTSS (Lin et al., 2019), GrPPG (Li et al., 2016) and LrPPG (Liu et al., 2016). The reason for this phenomenon is that the noise characteristics and heart rate in rPPG methods share a similar frequency range (Wang et al., 2016). Thus, the extracted pulse signal can be easily affected by facial expressions, illumination, and camera parameters (Farrukh, Aburas, Cao, & Wang, 2020; Kotwal et al., 2019). However, some rPPG based methods provide slightly better performance than our method, $PATRON$, like CFrPPG (Liu, Lan et al., 2018). However, it is important to understand that rPPG methods and respiratory signals utilize different frequency bands; thus, they provide different information that can be simultaneously utilized for improving anti-spoofing. In the future, we will be looking to improve anti-spoofing by consolidating the information provided by respiratory signal and cardiovascular pulse signal.

Upon rigorous analysis, we found that there are few samples that are incorrectly predicted by our proposed method but correctly classified by state-of-the-art methods. Similarly, few samples are correctly classified by the proposed method but incorrectly classified by the state-of-the-art methods. It is clearly depicted in Fig. 5. Our proposed method provides incorrect prediction when the input video contains significant face movement. These movements cause spurious tracking of discriminatory facial points and noisy temporal signals, resulting in less correlation of temporal signals. This behavior is depicted in Fig. 5, where we analyze the predicted samples and found that the incorrectly predicted samples have less correlated temporal signals.

Our proposed method estimates the respiratory signal from a face video for predicting the 3D face mask. In contrast, other state-of-the-art rPPG based methods utilize the heart rate information. Thus, our method and rPPG based methods employ different frequency ranges and noise characteristics. Hence, it can be inferred that the proposed method and other state-of-the-art methods work on different principles. We perform the statistical hypothesis testing between our proposed method and the other well-known methods for rigorously analyzing this aspect. In essence, we perform the statistical chi-square testing (Pandis, 2016) of our proposed method PATRON with other methods. The *p-values* of the our proposed method $PATRON$ and MS-LBP (Erdogmus & Marcel, 2014), CTA (Boulkenafet et al., 2015), sCNN-LE (Qu et al., 2019; Shao et al., 2018), MS-LTSS (Lin et al., 2019), GrPPG (Li et al., 2016), LrPPG (Liu et al., 2016), CFrPPG (Liu, Lan et al., 2018) are 0.031, 0.002, 0.026, 0.030, 0.042, 0.007, 0.041, 0.028 respectively. Since the *p-values* between our proposed method with other methods lies below 0.05, it can be deduced that our proposed method and other methods are independent.

4.5. Ablation study

Our proposed method, $PATRON$, utilizes several similarity features (namely, SF_1, SF_2, and SF_3), proper ROI selection, and the Lagrangian



(a) R_1 (b) R_2 (c) R_3

Fig. 6. Examples of different ROIs used in ablation study.

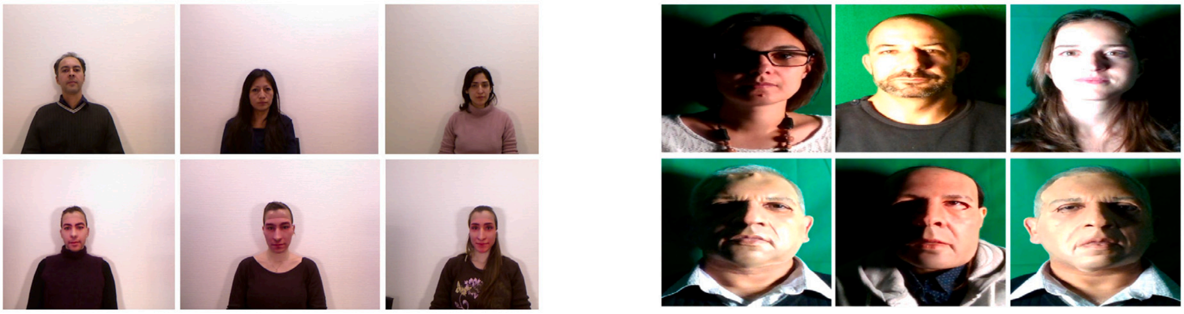
Table 3
Classifier selection of our proposed method, $PATRON$.

	AUC	EER
SVM (Linear)	77.9	24.8
SVM (Polynomial)	81.5	21.5
QDA	84.8	18.1
LDA	85.0	17.5
Gaussian NB	82.5	20.5
KNN	79.5	24.0
$PATRON$	**87.8**	**14.7**

approach for temporal signal extraction while avoids face alignment for 3D face mask anti-spoofing. In this subsection, we conduct experiments to analyze the importance of each of these aspects rigorously. For understanding the importance of proper ROI selection in the 3D face mask anti-spoofing methods, our additional experiments are conducted by utilizing the three different ROIs. Since the person performs respiration through the nose, the respiratory signal is prominent in the nose area (Bennett et al., 2017). Thus, the nose region is included in all the ROIs. The different ROIs utilized in the experiments are shown in Fig. 6. The R_1 is the ROI used by our method, $PATRON$. It covers the nose region along with some cheek regions but excludes the forehead, eyes, and lower mouth regions (refer Section 3.1 for more details). Likewise, the full nose region and nose tip region are selected in ROI R_2 and R_3, respectively. To this end, ROI R_2 is given by computing the minimum enclosing area rectangle (Boltz et al., 2009) of the following landmark points: (i) 29 of the nose; and (ii) 32 and 36 of the nose. Similarly, ROI R_3 is given by computing the minimum enclosing area rectangle (Boltz et al., 2009) of the following landmark points: (i) 30, 32, and 36 of the nose; and (ii) 52 of the upper lip.

To thoroughly analyzing of the importance of these components in $PATRON$, we compare the $PATRON$ method in Table 2 with other methods that are designed by (i) utilizing different combinations of similarity features; (ii) utilizing different ROI selection; (iii) utilizing or including face alignment; and (iv) utilizing either Eulerian approach or Lagrangian Approach. The details of these new methods are provided in Table 2. The methods based on directional similarity where the Eulerian approach cannot be used marked as "NA" in the table. To estimate the respiratory signal from a face video, we utilized the well-known method (Gupta et al., 2018b). It can be observed from the table that:

1. Out of all the three ROIs considered in the experiments, the best performance can be achieved when R_1 is selected as ROI while R_3 demonstrates the lowest performance. The ROI R_3 includes the facial region below the nose and above the lips, and this region is easily affected by the facial region. Due to this reason, performance degradation can be observed when ROI R_3 is considered. In contrast, ROIs R_1 and R_2 utilizes facial regions which are least affected by facial expressions. Furthermore, ROI R_1 provide better efficacy than ROI R_3 because it also includes the left and right cheek regions, which: (i) contains relevant respiratory signal information (Al-Khalidi, Saatchi, Burke, & Elphick, 2010; Kwon, Kim, Lee, & Park, 2015); (ii) are less affected by facial expressions; and (iii) provides more distinctive facial points.
2. It can be observed from the table that there is not a particular similarity feature that provides the best performance in all the scenarios, that is, irrespective of the selection of ROI, face registration, and temporal signal extraction. However, it can be observed that the performance increase when different similarity features are consolidated, and the best performance can be achieved by consolidating all the proposed similarity features, which are SF_1, SF_2, and SF_3. It indicates that our similarity features provide some complementary information.
3. It can be seen from the table that the Eulerian approach performs slightly better than the Lagrangian approach when R_2 and R_3 are used as ROI rather than R_1. The reason for lower performance in R_2 and R_3 is that usually, there are fewer distinctive points available in R_2 and R_3 for tracking, as compared to R_1. However, when a larger number of distinctive points are available, the Lagrangian approach outperforms the Eulerian approach because tracking the temporal movement of distinctive facial points using the Lagrangian approach is less erroneous than analyzing the temporal color variations in the Eulerian approach.
4. The face alignment has been used extensively in the literature to improve the efficacy (Gupta et al., 2020). Hence, we performed some experiments to evaluate its effectiveness. Ironically, it degrades the performance rather than improving it. Face alignment aims to minimize the rigid deformations in the face and thereby stabilizes the face (Gupta et al., 2018b). The state-of-the-art face alignment method (Zadeh et al., 2017) performs the alignment by detecting the eye centers and normalizes them to a fixed location. Unfortunately, the eye centers cannot be correctly estimated in the case of eye blinking. It subsequently results in erroneous face alignment whenever there is inevitable eye blinking (Gupta et al., 2020). Due to this reason, one can observe significant performance degradation when face alignment is utilized.



(a) 3DMAD

(b) CSMAD

Fig. 7. Example of face images from the 3DMAD and CSMAD datasets. The first and second row images belong to genuine and spoofed videos, respectively.

5. The best performance can be achieved when all the similarity features extracted from R_1 ROI using the Lagrangian approach while avoiding face alignment are consolidated. The performance is degraded when any of its conditions are violated. Thus, these conditions are used in our method, $PATRON$, to achieve the best possible performance.

We also study the efficacy of different classifiers for differentiating genuine and spoof videos. Table 3 demonstrate the performance of well-known classifiers, namely Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), SVM with Linear kernel, SVM with Polynomial kernel of degree 3, SVM using RBF kernel, k-nearest neighbors (KNN), and Gaussian Naive Bayes (Gaussian NB). In KNN, we try KNN for different values of k, and the best results are achieved when k is set to 3. The corresponding results are shown in the table. It can be observed from the table that SVM with RBF kernel used by the proposed method $PATRON$ performs better than the other considered classifiers. Usually, classifiers like LDA, QDA and SVM with Linear kernel are helpful when the input features are linearly separable. However, SVM with RBF kernel is useful for non-linearly separable features. In our case, it performs better than the considered linearly separable classifiers. Thus, it can be inferred that our similarity features are not linearly separable. Furthermore, our similarity features are dependent on each other because they analyze the respiratory signals in different ways. Due to this reason, Gaussian NB, which is useful when features are conditionally independent, is insufficient in our case (refer Table 3). However, the reason for the poor efficacy of KNN in our case can be attributed to the limited training data.

5. Discussion

In this paper, we performed extensive experiments on the publicly available dataset HKBU-MARs-V1+ and justified the importance of our ROI selection and similarity features for 3D face mask detection. Apart from the HKBU-MARs-V1+ dataset, there exist two more publicly available datasets for 3D face mask detection, which are 3DMAD (Erdogmus & Marcel, 2014) and CSMAD (Bhattacharjee, Mohammadi, & Marcel, 2018). However, we have not considered these datasets. The 3DMAD dataset is avoided because it consists of the face video having a small face resolution of around 80 × 80 pixels. In the respiratory signal, ROI is extracted only around the nose region. Such a small face resolution is insufficient to extract the ROI for the respiratory signal. Even in most of the samples, we do not get any ROI. On the other hand, the reason for avoiding the CSMAD dataset is that it contains face videos that are highly affected by the low illumination. In some videos, the face is not visible even to the human eyes, and the state-of-the-art face detector, OpenFace (Zadeh et al., 2017). These problems can be visualized from the samples of 3DMAD and CSMAD datasets, which are shown in Fig. 7. Furthermore, the HKBU-MARsV1+ dataset considered in the experiments contains more variations than 3DMAD and CSMAD. That is, 3DMAD and CSMAD utilize only one type of mask, whereas HKBU-MARsV1+ uses two different types of mask and thus, offers more challenging 3D face mask detection settings.

Our proposed method, $PATRON$, compactly represents the correlation between the temporal signals, respiratory signal, and direction signals using statistical features. If respiratory signals are used to train the classifier, then we achieved the performance in terms of AUC and EER as 72.4% and 36.8%, respectively. Thus, performance degradation can be observed when respiratory signals are used instead of compact similarity features. The reason for this behavior is that the classifier requires a large number of parameters learning with limited training data when the respiratory signal is utilized and the amplitude of respiratory signal varies with a different subject.

6. Conclusion

Face authentication is avoided in safety-critical applications because an adversary can easily spoof it. The existing face anti-spoofing methodologies are based on motion, appearance, and rPPG. These methodologies and their consolidation are insufficient to detect the 3D face mask attack, which is considered the most potent attack. Thus, this paper has explored a different methodology for face anti-spoofing. It utilized the respiration rate, which provides complementary information with the existing methodologies. Our experimental results conducted on a highly challenging dataset, HKBU-MARsV1+, have justified that respiratory signal can be considered for 3D face mask anti-spoofing. The results have also indicated that our respiration-based method can perform similar to several state-of-the-art systems. However, it requires that the respiratory signal is properly extracted from relevant facial regions and the appropriate features are utilized for the anti-spoofing. To this end, our compact novel similarity features played a crucial role. Furthermore, the results have demonstrated that the performance of 3D face mask anti-spoofing using respiration rate can be significantly degraded due to temporal signal extraction (whether Lagrangian or Eulerian approach) and face alignment. Hence, both these factors should be thoroughly evaluated. In addition, it is important to remember that our proposed method, $PATRON$ and existing rPPG methods are based on contradictory principles. In the future, we will be looking to consolidate the $PATRON$ with existing rPPG methods to improve the efficacy of 3D face mask anti-spoofing further.

CRediT authorship contribution statement

Lokendra Birla: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Puneet Gupta:** Conceptualization, Methodology, Validation, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Supervision, Project administration.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

The authors would like to thank all those researchers who have provided access to the HKBU 3D Mask Attack with Real-world-variations (HKBU-MARs) (Li et al., 2016) and 3DMAD (Erdogmus & Marcel, 2014) datasets.

References

Al-Khalidi, F. Q., Saatchi, R., Burke, D., & Elphick, H. (2010). Tracking human face features in thermal images for respiration monitoring. In *IEEE international conference on computer systems and applications (ICCSA)* (pp. 1–6). IEEE.

Atoum, Y., Liu, Y., Jourabloo, A., & Liu, X. (2017). Face anti-spoofing using patch and depth-based CNNs. In *IEEE international joint conference on biometrics (IJCB)* (pp. 319–328).

Balakrishnan, G., Durand, F., & Guttag, J. (2013). Detecting pulse from head motions in video. In *IEEE conference on computer vision and pattern recognition (CVPR)* (pp. 3430–3437).

Bao, W., Li, H., Li, N., & Jiang, W. (2009). A liveness detection method for face recognition based on optical flow field. In *International conference on image analysis and signal processing (ICIASP)* (pp. 233–236). IEEE.

Bennett, S. L., Goubran, R., & Knoefel, F. (2017). Comparison of motion-based analysis to thermal-based analysis of thermal video in the extraction of respiration patterns. In *International conference of the IEEE engineering in medicine and biology society (EMBC)* (pp. 3835–3839). IEEE.

Bhattacharjee, S., Mohammadi, A., & Marcel, S. (2018). Spoofing deep face recognition with custom silicone masks. In *IEEE international conference on biometrics theory, applications and systems (BTAS)* (pp. 1–7). IEEE.

Boltz, S., Debreuve, E., & Barlaud, M. (2009). High-dimensional statistical measure for region-of-interest tracking. *IEEE Transactions on Image Processing*, *18*(6), 1266–1283.

Boulkenafet, Z., Komulainen, J., & Hadid, A. (2015). Face anti-spoofing based on color texture analysis. In *IEEE international conference on image processing (ICIP)* (pp. 2636–2640). IEEE.

Boulkenafet, Z., Komulainen, J., & Hadid, A. (2016a). Face antispoofing using speeded-up robust features and fisher vector encoding. *IEEE Signal Processing Letters*, *24*, 141–145.

Boulkenafet, Z., Komulainen, J., & Hadid, A. (2016b). Face spoofing detection using colour texture analysis. *IEEE Transactions on Information Forensics and Security*, *11*(8), 1818–1830.

Cattani, L., Alinovi, D., Ferrari, G., Raheli, R., Pavlidis, E., Spagnoli, C., et al. (2017). Monitoring infants by automatic video processing: A unified approach to motion analysis. *Computers in Biology and Medicine*, *80*, 158–165.

Chen, M., Zhu, Q., Zhang, H., Wu, M., & Wang, Q. (2019). Respiratory rate estimation from face videos. In *IEEE EMBS international conference on biomedical & health informatics* (pp. 1–4).

Chingovska, I., Anjos, A., & Marcel, S. (2012). On the effectiveness of local binary patterns in face anti-spoofing. In *International conference of biometrics special interest group (BIOSIG)* (pp. 1–7).

Erdogmus, N., & Marcel, S. (2014). Spoofing face recognition with 3D masks. *IEEE Transactions on Information Forensics and Security*, *9*, 1084–1097.

Farrukh, H., Aburas, R. M., Cao, S., & Wang, H. (2020). FaceRevelio: a face liveness detection system for smartphones with a single front camera. In *International conference on mobile computing and networking (MobiCom)* (pp. 1–13).

Feng, L., Po, L.-M., Li, Y., Xu, X., Yuan, F., Cheung, T. C.-H., et al. (2016). Integration of image quality and motion cues for face anti-spoofing: A neural network approach. *Journal of Visual Communication and Image Representation*, *38*, 451–460.

de Freitas Pereira, T., Komulainen, J., Anjos, A., De Martino, J. M., Hadid, A., Pietikainen, M., et al. (2014). Face liveness detection using dynamic texture. *EURASIP Journal on Image and Video Processing*, *2014*, 2.

Guo, Y., Liu, X., Peng, S., Jiang, X., Xu, K., Chen, C., et al. (2020). A review of wearable and unobtrusive sensing technologies for chronic disease management. *Computers in Biology and Medicine*, Article 104163.

Gupta, P., Bhowmick, B., & Pal, A. (2017a). Accurate heart-rate estimation from face videos using quality-based fusion. In *IEEE international conference on image processing (ICIP)* (pp. 4132–4136).

Gupta, P., Bhowmick, B., & Pal, A. (2017b). Serial fusion of Eulerian and Lagrangian approaches for accurate heart-rate estimation using face videos. In *Annual international conference of the IEEE engineering in medicine and biology society (EMBC)* (pp. 2834–2837). IEEE.

Gupta, P., Bhowmick, B., & Pal, A. (2018a). Exploring the feasibility of face video based instantaneous heart-rate for micro-expression spotting. In *IEEE conference on computer vision and pattern recognition workshops (CVPRW)* (pp. 1316–1323).

Gupta, P., Bhowmick, B., & Pal, A. (2020). MOMBAT: Heart rate monitoring from face video using pulse modeling and Bayesian tracking. *Computers in Biology and Medicine*, Article 103813.

Gupta, P., Bhowmik, B., & Pal, A. (2018b). Robust adaptive heart-rate monitoring using face videos. In *IEEE winter conference on applications of computer vision (WACV)* (pp. 530–538).

Gupta, P., & Gupta, P. (2015). An accurate finger vein based verification system. *Digital Signal Processing*, *38*, 43–52.

Gupta, P., & Gupta, P. (2016). An accurate fingerprint orientation modeling algorithm. *Applied Mathematical Modelling*, *40*(15–16), 7182–7194.

Hamedani, K., Bahmani, Z., & Mohammadian, A. (2016). Spatio-temporal filtering of thermal video sequences for heart rate estimation. *Expert Systems with Applications*, *54*, 88–94.

Hirsch, G., Jensen, S. H., Poulsen, E. S., & Puthusserypady, S. (2020). Atrial fibrillation detection using heart rate variability and atrial activity: A hybrid approach. *Expert Systems with Applications*, Article 114452.

Jourabloo, A., Liu, Y., & Liu, X. (2018). Face de-spoofing: Anti-spoofing via noise modeling. In *European conference on computer vision (ECCV)* (pp. 290–306).

Kalal, Z., Mikolajczyk, K., & Matas, J. (2010). Forward-backward error: Automatic detection of tracking failures. In *International conference on pattern recognition (ICPR)* (pp. 2756–2759).

Kollreider, K., Fronthaler, H., & Bigun, J. (2008). Verifying liveness by multiple experts in face biometrics. In *IEEE conference on computer vision and pattern recognition workshops (CVPRW)* (pp. 1–6).

Kollreider, K., Fronthaler, H., Faraj, M. I., & Bigun, J. (2007). Real-time face detection and motion analysis with application in "liveness" assessment. *IEEE Transactions on Information Forensics and Security*, *2*, 548–558.

Kotwal, K., Bhattacharjee, S., & Marcel, S. (2019). Multispectral deep embeddings as a countermeasure to custom silicone mask presentation attacks. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, *1*, 238–251.

Kwon, S., Kim, J., Lee, D., & Park, K. (2015). ROI analysis for remote photoplethysmography on facial video. In *Annual international conference of the IEEE engineering in medicine and biology society (EMBC)* (pp. 4938–4941). IEEE.

Li, X., Komulainen, J., Zhao, G., Yuen, P.-C., & Pietikainen, M. (2016). Generalized face anti-spoofing by detecting pulse from face videos. In *International conference on pattern recognition (ICPR)* (pp. 4244–4249).

Li, J., Wang, Y., Tan, T., & Jain, A. K. (2004). Live face detection based on the analysis of fourier spectra. In *Biometric technology for human identification, Vol. 5404* (pp. 296–303). International Society for Optics and Photonics.

Lin, B., Li, X., Yu, Z., & Zhao, G. (2019). Face liveness detection by r-PPG features and contextual patch-based cnn. In *International conference on biometric engineering and applications (ICBEA)* (pp. 61–68).

Liu, Y., Jourabloo, A., & Liu, X. (2018). Learning deep models for face anti-spoofing: Binary or auxiliary supervision. In *IEEE conference on computer vision and pattern recognition (CVPR)* (pp. 389–398).

Liu, S.-Q., Lan, X., & Yuen, P. C. (2018). Remote photoplethysmography correspondence feature for 3D mask face presentation attack detection. In *European conference on computer vision (ECCV)* (pp. 558–573).

Liu, S., Lan, X., & Yuen, P. (2020). Temporal similarity analysis of remote photoplethysmography for fast 3D mask face presentation attack detection. In *IEEE winter conference on applications of computer vision (WACV)* (pp. 2608–2616).

Liu, Y., Tai, Y., Li, J., Ding, S., Wang, C., Huang, F., et al. (2019). Aurora guard: Real-time face anti-spoofing via light reflection. arXiv preprint arXiv:1902.10311.

Liu, S., Yuen, P. C., Zhang, S., & Zhao, G. (2016). 3D mask face anti-spoofing with remote photoplethysmography. In *European conference on computer vision (ECCV)* (pp. 85–100). Springer.

Luguern, D., Perche, S., Benezeth, Y., Moser, V., Andrea Dunbar, L., Braun, F., et al. (2020). An assessment of algorithms to estimate respiratory rate from the remote photoplethysmogram. In *IEEE conference on computer vision and pattern recognition workshops (CVPRW)* (pp. 304–305).

Maatta, J., Hadid, A., & Pietikinen, M. (2011). Face spoofing detection from single images using micro-texture analysis. In *International joint conference on biometrics (IJCB)* (pp. 1–7).

Nowara, E. M., Sabharwal, A., & Veeraraghavan, A. (2017). PPGSecure: Biometric presentation attack detection using photopletysmograms. In *IEEE international conference on automatic face & gesture recognition (FG)* (pp. 56–62).

Pandis, N. (2016). The chi-square test. *American Journal of Orthodontics and Dentofacial Orthopedics*, *150*(5), 898–899.

Patel, K., Han, H., Jain, A. K., & Ott, G. (2015). Live face video vs. spoof face video: Use of moire patterns to detect replay video attacks. In *International conference on biometrics (ICB)* (pp. 98–105).

Qu, X., Dong, J., & Niu, S. (2019). ShallowCNN-LE: A shallow CNN with Laplacian embedding for face anti-spoofing. In *IEEE international conference on automatic face & gesture recognition (FG)* (pp. 1–8). IEEE.

Rehman, Y. A. U., Po, L. M., & Liu, M. (2018). LiveNet: Improving features generalization for face liveness detection using convolution neural networks. *Expert Systems with Applications*, *108*, 159–169.

Rehman, Y. A. U., Po, L.-M., & Liu, M. (2020). SLNet: Stereo face liveness detection via dynamic disparity-maps and convolutional neural network. *Expert Systems with Applications*, *142*, Article 113002.

Sanyal, S., & Nundy, K. K. (2018). Algorithms for monitoring heart rate and respiratory rate from the video of a user's face. *IEEE Journal of Translational Engineering in Health and Medicine*, *6*, 1–11.

Shao, R., Lan, X., & Yuen, P. C. (2018). Joint discriminative learning of deep dynamic textures for 3D mask face anti-spoofing. *IEEE Transactions on Information Forensics and Security*, *14*, 923–938.

Shi, J., et al. (1994). Good features to track. In *IEEE conference on computer vision and pattern recognition (CVPR)* (pp. 593–600).

Wang, W., den Brinker, A. C., Stuijk, S., & de Haan, G. (2016). Algorithmic principles of remote PPG. *IEEE Transactions on Biomedical Engineering*, *64*(7), 1479–1491.

Wang, Y., Nian, F., Li, T., Meng, Z., & Wang, K. (2017). Robust face anti-spoofing with depth information. *Journal of Visual Communication and Image Representation*, *49*, 332–337.

Wen, D., Han, H., & Jain, A. K. (2015). Face spoof detection with image distortion analysis. *IEEE Transactions on Information Forensics and Security*, *10*, 746–761.

Yan, J., Zhang, Z., Lei, Z., Yi, D., & Li, S. Z. (2012). Face liveness detection by exploring multiple scenic clues. In *International conference on control automation robotics & vision (ICARCV)* (pp. 188–193). IEEE.

Zadeh, A., Chong Lim, Y., Baltrusaitis, T., & Morency, L.-P. (2017). Convolutional experts constrained local model for 3D facial landmark detection. In *IEEE international conference on computer vision workshops (ICCVW)* (pp. 2519–2528).