3D Mask Face Anti-spoofing with Remote Photoplethysmography

Siqi Liu[1], Pong C. Yuen[1(✉)], Shengping Zhang[2], and Guoying Zhao[3]

[1] Department of Computer Science,
Hong Kong Baptist University, Kowloon Tong, Hong Kong
{siqiliu,pcyuen}@comp.hkbu.edu.hk
[2] School of Computer Science and Technology,
Harbin Institute of Technology, Harbin, China
s.zhang@hit.edu.cn
[3] Center for Machine Vision and Signal Analysis,
University of Oulu, Oulu, Finland
gyzhao@ee.oulu.fi

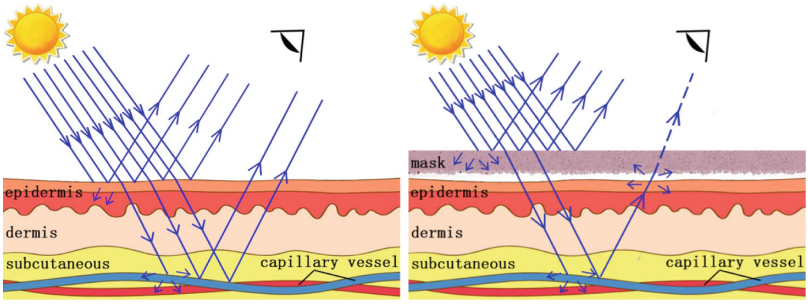
Abstract. 3D mask spoofing attack has been one of the main challenges in face recognition. Among existing methods, texture-based approaches show powerful abilities and achieve encouraging results on 3D mask face anti-spoofing. However, these approaches may not be robust enough in application scenarios and could fail to detect imposters with hyper-real masks. In this paper, we propose a novel approach to 3D mask face anti-spoofing from a new perspective, by analysing heartbeat signal through remote Photoplethysmography (rPPG). We develop a novel local rPPG correlation model to extract discriminative local heartbeat signal patterns so that an imposter can better be detected regardless of the material and quality of the mask. To further exploit the characteristic of rPPG distribution on real faces, we learn a confidence map through heartbeat signal strength to weight local rPPG correlation pattern for classification. Experiments on both public and self-collected datasets validate that the proposed method achieves promising results under intra and cross dataset scenario.

Keywords: Face anti-spoofing · 3D mask attack · Remote photoplethysmography

1 Introduction

Face recognition has been widely employed in a variety of applications. Like any other biometric modality [1,2], a critical concern in face recognition is to detect spoofing attack. In the past decade, photos and videos are two popular media of carrying out spoofing attacks and varieties of face anti-spoofing algorithms have been proposed [1–12] and encouraging results have been obtained. Recently, with the rapid development of 3D reconstruction and material technologies, 3D mask attack becomes a new challenge to face recognition since affordable off-the-shelf

B. Leibe et al. (Eds.): ECCV 2016, Part VII, LNCS 9911, pp. 85–100, 2016.
DOI: 10.1007/978-3-319-46478-7_6



(a) Genuine Face (b) Masked Face

Fig. 1. Effect of remote photoplethysmography (rPPG) on normal unmasked face (a), and masked face (b). (a) shows rPPG on a genuine face: Sufficient light penetrate the semi-transparent skin tissue and interact with blood vessels. rPPG signal can go through skin and be detected by RGB camera. (b) depicts rPPG on a masked face: The mask material blocks large portion of the light that the skin should absorb. Light source needs to penetrate a layer of painted plastic and a layer of skin before interacting with the blood. Remain rPPG signals will be too weak to be detected

masks[1] have been shown to be able to spoof existing face recognition system [13]. Unlike the success in traditional photo or video based face anti-spoofing, very few methods have been proposed to address 3D mask face anti-spoofing. To the best of our knowledge, most existing face anti-spoofing methods are not able to tackle this new attack since 3D masks have similar appearance and geometry properties as live faces.

Texture-based methods are the few effective approaches that has been evaluated on 3D mask attack problem [13]. Experimental results demonstrate their strong discriminative ability on 3DMAD and Morpho datasets with different classifiers [13]. Through the concatenation of different LBP settings, Multi-Scale LBP can effectively capture the subtle texture difference between genuine and masked faces and achieves 99.4 % AUC on 3DMAD dataset [13]. Although the results are promising, the problem of the cross-dataset (where training and testing data are selected from different datasets) scenario remains open. From the application perspective, it is essential for a face anti-spoofing method to be effective and robust to different mask types and video qualities. In fact, as reported in [14,15], texture based methods mentioned in [9,13,16] cannot be well generalized under inter-test (cross-dataset) protocol [14]. This is because of the over-fitting problem due to its intrinsic data-driven nature [15]. Also, since the texture-based methods rely on the appearance difference between 3D masks and genuine faces, it may not work for the super realistic masks that have imperceptible difference with the genuine face, e.g., prosthetics makeup.

To address the aforementioned limitations, we propose a novel approach to 3D mask face anti-spoofing from a new perspective, by using heart rate signal

[1] www.thatsmyface.com.

as a more intrinsic cue for mask detection. Photoplethysmography (PPG), as one of the general ways for heart rate monitoring, could be used to detect this intrinsic liveness information. However, we can hardly adapt it into existing systems since PPG extracts heartbeat from color variation of blood through pulse oximeter in an contact way. In recent years, based on the same principle, researches find that the vital signal can be detected remotely through web-camera [17]. This new technique is named as remote Photoplethysmography (rPPG) [18]. Due to the non-contact property, rPPG could be a possible solution for the 3D mask face anti-spoofing problem [19]. The principle is presented in Fig. 1. rPPG detects vascular blood flow based on the absorption and reflection of light passing through human skin. For a genuine face, although part of the light is reflected or absorbed by the semi-transparent human skin, heartbeat signal can still be detected from the subtle blood color variation. For a masked face in Fig. 1(b), the light source needs to penetrate a layer of painted plastic and a layer of skin before interacting with the blood. Such a small amount of energy results in a very noisy rPPG signal, if not impossible, to detect the blood volume flow.

Based on this principle, we propose to use rPPG for 3D mask face anti-spoofing. An intuitive solution is to extract the global heartbeat signal through rPPG from face video as the vital sign. Theoretically, heartbeat should show high amplitude on a genuine face and very low amplitude on a masked face. However, the global method may not be able to achieve good performance since interference like poor video quality, low exposure condition, light change or head motion may conceal the subtle heart rate signal and introduce false rejection error (see Sect. 3 for detailed analysis). Moreover, the global solution lacks spatial information which may lead false accept error since rPPG signal may still be obtained on partially masked face. As such, we propose to use rPPG from local perspective. Existing studies indicate that rPPG signal strength varies along local face region [20]. Forehead and cheek with dense capillary vessels can provide stronger and clearer rPPG signals than other areas. Meanwhile, based on our observation, the local rPPG strength forms a stable spatial pattern along different subjects. Therefore, the local rPPG signals could be used to form a discriminative pattern for 3D mask detection.

In summary, the contributions of this paper are listed below:

- We propose to use face rPPG signals as the natural and intrinsic sign for 3D mask face anti-spoofing, which would perform well regardless of mask appearance quality.
- We develop a novel local rPPG-based face anti-spoofing method to model the face heart rate pattern through the cross-correlation of local rPPG signals. With the confidence evaluation of local signals, the genuine faces can be differentiated from the masked faces effectively.

The organization of this paper is as follows. We review the related work in Sect. 2. Then, the principle analysis of our local rPPG-based solution is given in Sect. 3. After that, we describe the proposed method in Sect. 4 and report the

experimental results in Sect. 5. Finally, we conclude this paper by drawing a few remarks in Sect. 6.

2 Related Work

2.1 Face Anti-spoofing

Existing face anti-spoofing methods can be mainly divided into two categories: appearance based approaches and motion based approaches. As the appearance of the printed photos and videos may differ from the real faces, texture-based approaches have been used to detect printed or displayed artifacts and achieve encouraging results [5,9,12]. Multi-Scale [5] LBP concatenates different LBP settings and achieves promising performance on 3D mask detection [13]. While the results are promising in the above methods, recent studies indicate that they cannot generalize well in the cross-dataset scenario [14,15]. Deep learning based methods [21] also achieve encouraging results on 3DMAD. But they may also face the same problem due to the intrinsic data-driven nature. Image distortion analysis (IDA) based approaches perform well in the cross-dataset scenario [15]. But for 3D mask attack, these methods may not stand as the masked face has no relation to the video or image quality.

Motion-based approaches use unconscious face motion or human-computer interaction (HCI) to detect photo and video attacks through user's response (e.g., detect whether the user blinks unconsciously or being instructed to do so [8,9, 22]). These approaches are particularly effective against photo and stationary screen attacks. However, when facing mask attack exposes eyes or mouth, or video attack contents face motion, they may not work effectively.

There are also other approaches based on different cues, which achieve desired performance under various assumptions [11,23,24]. For example, [24] solves the problem through spoofing medium shape (context). These methods may not be able to tackle the mask attack since 3D mask faces have the same geometric property as real faces. Multi-spectrum analysis may work since it relies on the fact that the frequency responses of 3D mask faces and real faces are different. However, it requires specific equipments to capture the invisible light which may not be economical for a face recognition system.

2.2 Remote Photoplethysmography

rPPG is a new research topic in medical field and only few methods are proposed in recent years. Verkruysse et al. [17] is one of the early work that evaluates rPPG under ambient light. Poh et al. [25] and Lewandowska et al. [26] propose to use blind source separation (BSS) techniques, e.g., independent component analysis (ICA) and principle component analysis (PCA), to extract rPPG signals from a face video. Lempe et al. [20] observes that the variation of rPPG is sensitive to different facial parts. de Haan and Jeanne [18] models the physical process of rPPG to achieve motion robustness. Li et al. [19] builds a framework

that contains illumination rectification and motion elimination to achieve good performance in realistic situations. Recently, matrix completion technique is also applied to achieve better robustness [27].

3 Why Does Local rPPG Work for 3D Mask Face Anti-spoofing?

In this section, we explain the reasons why local rPPG works for 3D mask face anti-spoofing. We first analyse the principle of rPPG signals from live face and mask, respectively and then demonstrate why local rPPG is effective for mask attack.

3.1 Analysis of rPPG Signal on Live and Masked Face

As shown in Fig. 1(a), light illuminates capillary vessel and rPPG signal penetrates skin to be observed. Thus, the observed signal from a live face $\hat{s}_l$ can be modeled as follows,

$$\hat{s}_l = T_s I s + \epsilon \tag{1}$$

where s is the raw rPPG signal from capillary vessels, T_s is the transmittance of skin and I is the mean intensity of facial skin under ambient light. ϵ is the environmental noise.

For a masked face shown in Fig. 1(b), the light need to go through the mask before interacting with capillaries. Also, source rPPG signal need to penetrate the mask before captured by camera. So, the observed signal $\hat{s}_m$ can be represented as

$$\hat{s}_m = T_m T_s I_m s + \epsilon$$

where T_m is the transmittance of mask and I_m is the mean intensity of face under mask. I_m can be modeled as $I_m = T_m I$. With simple deduction, the observed signal from the masked face can be represented as

$$\begin{aligned} \hat{s}_m &= T_m^2 T_s I s + \epsilon \\ &= T_m^2 \hat{s}_l + \epsilon \end{aligned} \tag{2}$$

Considering the transmittance of existing mask material, the rPPG signal from a masked face is too weak to be detected, which leads to the feasibility of our proposed method. Hence, rPPG signal can be detected on genuine face, but not masked face.

3.2 Local rPPG for 3D Mask Face Anti-spoofing

Based on the analysis in Sect. 3.1, 3D masked faces can be distinguished from real faces by analysing rPPG signals extracted from the global face. Unfortunately, the global rPPG signals could be too weak to be detected in real application scenario. From Eq. 1, $\hat{s}_l$ is proportional to the intensity I. As shown in Fig. 1(a),

rPPG signals are weak (around ±2 variations for a 8-bit color camera [28]) since only a small portion of light can transmit to blood vessels as quite amount of light energy is reflected or absorbed by human skin. Hence, poor video quality such as inadequate exposure will weaken $\hat{s}_l$ and increase the difficulty of detection. Also, rPPG is sensitive to illumination change since it is based on subtle heartbeat-related color variation of the ROI during a specific time interval. Face motion may also conceal the rPPG signal by introducing imprecise tracking or skin angle change [18]. Meanwhile, when a subject is under single light source, head motion may also cause intensity changes on face. This is because facial structure, e.g., hair or nose, will cast shadow on skin region and motion will change its area thereby influence the intensity. Therefore, we can conclude that many interference like video quality, light change or head motion may conceal the subtle heart rate signal. In other words, false rejection error will be made since we may not be able to detect vital sign on genuine face. Moreover, for partially covered mask, vital signal can still be obtained from the exposed part, such as cheek and forehead [17], which may contribute strong heart rate signal and be regarded as a liveness evidence which leads to the failure on face anti-spoofing. As such, even if global rPPG signal is detected from a subject, we cannot directly regard the one as a genuine face.

In sum, we propose to adopt local rPPG signals for 3D mask face anti-spoofing. Existing studies indicate that strength of rPPG signals vary along local face regions [17,20]. Flat regions such as forehead and cheek with dense capillary vessels can provide stronger and clearer rPPG signals than other areas. Also, through observation of numbers of subjects, we found that the local rPPG signal strength forms a stable pattern for different people. In other words, the local rPPG signals could be used to form a discriminative and robust pattern for 3D mask detection.

4 Proposed Method

Based on the analysis in Sect. 3, we propose a novel 3D mask face anti-spoofing approach by exploiting the characteristic of local rPPG extracted from 3D mask faces and real faces.

4.1 Overview

The overview of the proposed method is presented in Fig. 2, which contains four main components: (1) local rPPG extraction, (2) local rPPG correlation modeling, (3) confidence map learning and (4) classification. First, to avoid imperfect boundary from facial motion, face landmarks are detected [29] so as to divide a face into a number of local regions (see Sect. 5.1 for implementation details). Then, local rPPG signals are extracted from these local face regions. To make the extracted rPPG signals robust to head motion and noise, we adopt de Haan and Jeanne method [18] as the rPPG sensor on local face regions. In training stage, the local heartbeat signal patterns are extracted through the proposed

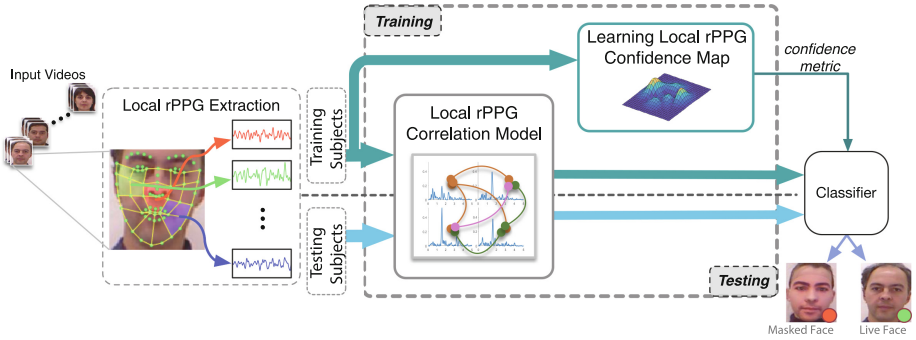


Fig. 2. Block diagram of the proposed method. Four main components are included: (1) local rPPG extraction, (2) local rPPG correlation modeling, (3) confidence map learning and (4) classification. From input face video, local rPPG signals are extracted from the local regions selected along landmarks. After that, the proposed local rPPG correlation model extract discriminative local heartbeat signal pattern through cross-correlation of input signals. In training stage, local rPPG confidence map is learned and transformed into metric to measure the local rPPG correlation pattern. Finally, local rPPG correlation pattern and confidence metric is fed into classifier.

local rPPG correlation model. At the same time, we use training subjects to learn the local rPPG confidence map and transform it into distance metric for classification. In testing stage, when a test face is presented to the system, local rPPG correlation features are also extracted from the testing subjects. Finally, result is obtained through the classification.

4.2 Local rPPG Correlation Model

Given the local rPPG signal $[\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_N]^\intercal$, we could model the local rPPG pattern by directly extracting the features of signal, such as the signal-to-noise ratio (SNR), maximum amplitude, or power spectrum density (PSD). Then, the final decision could be made by feeding the extracted features into a classifier. However, this intuitive model can not generalize well because of the following reasons: (1) The rPPG amplitude varies in different region with different people. The intuitive solution may not be able to adapt the signal amplitude variation along different subjects. (2) rPPG strength varies along video quality under cross-dataset scenario. It means the classifier may over-fit on high quality video contains clear rPPG signal. When encountering genuine testing samples from unseen low quality video, the vital sign may not be strong enough so that the classifier may regard it as mask.

Recall the rPPG principle is measuring human pulse rate through the blood flow variation caused by heart beat. It indicates that, for a sample subject, rPPG signals from different local regions should have similar shape with very small difference. To the best of our knowledge, this small retardation is likely because that the blood speed and vessels length from heart to local region has

small difference. It implies that, local rPPG signals should have great consistency on genuine face. While for masked face, they should have small frequency similarity and periodicity since the vital signals are blocked and the remaining signal mainly contains environmental noise. Therefore, we model the local rPPG pattern through the union of similarity of all the possible combination as follow:

$$\boldsymbol{x} = \bigcup_{\substack{i,j=1,\ldots,N \\ i \leqslant j}} \rho(\boldsymbol{s}_i, \boldsymbol{s}_j) \tag{3}$$

where $\rho(\boldsymbol{s}_i, \boldsymbol{s}_j)$ measures the similarity between two signals $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$, and the union $\bigcup$ is the concatenation operator. To measure the similarity between two signals with periodic frequencies, we define the similarity $\rho(\boldsymbol{s}_i, \boldsymbol{s}_j)$ as the maximum value of the cross-correlation spectrum of two local rPPG signals

$$\rho(\boldsymbol{s}_i, \boldsymbol{s}_j) = \max |\mathcal{F}\{\boldsymbol{s}_i \star \boldsymbol{s}_j\}| \tag{4}$$

where $\mathcal{F}$ is the Fourier transform and $\star$ is the cross-correlation operator. The resulting local rPPG correlation pattern is a $C(N, 2) + N$ dimensional feature.

Note that the signal $\boldsymbol{s}$ is not a feature vector. So we cannot simply using Euclidean distance to measure the similarity ρ between $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$. Thus, we design to simultaneously find out the periodicity and measure its frequency similarity. By doing the cross-correlation operation in Eq. 4, we could filter out the shared heartbeat related frequency and abate the random noise. Meanwhile, signals extracted from local masked face regions will suppress with each other because they are random noise and do not share the same periodic frequency. Therefore, 3D mask can be effectively detected since the local rPPG correlation pattern $\boldsymbol{x}$ will show a stable distribution on liveness face but not for masked face.

4.3 Learning Local rPPG Confidence Map

Given the local rPPG signal $[\boldsymbol{s}_1, \boldsymbol{s}_2, ...\boldsymbol{s}_N]$, the local rPPG correlation pattern can be discriminative under well controlled conditions. However, when encountering poor video quality, e.g., low exposure rate, the performance may drop since rPPG signals may be too weak and concealed by noise. Recall the principle (analysed in Sect. 3) that rPPG signal strength varies along local face region with a stable spatial distribution, we could boost the discriminative ability of $\boldsymbol{x}$ by emphasizing the robust regions which contain strong heartbeat signal and weaken the unreliable regions which contain less heartbeat signals or are fragile under interferences. To this end, we propose to learn the confidence map of local rPPG signals through the signal quality from training subjects.

Given J training subjects, considering a learning function y, which maps the signal quality to a real value, such that the average quality is maximized, i.e.,

$$\arg\max_{y} \sum_{j=1}^{J} y(g(\boldsymbol{s}^j, \boldsymbol{e}^j)) \tag{5}$$

where $g(\boldsymbol{s}^j, \boldsymbol{e}^j)$ measures the signal quality of $\boldsymbol{s}^j$ given its "ground truth" heart rate signal $\boldsymbol{e}^j$. As analysed in [28], the quality measure g can be defined by

$$g(\boldsymbol{s}^j, \boldsymbol{e}^j) = \frac{\sum_{f_{HR-r}}^{f_{HR+r}} \hat{\boldsymbol{s}}^j(f)}{\sum \hat{\boldsymbol{s}}^j(f) - \sum_{f_{HR-r}}^{f_{HR+r}} \hat{\boldsymbol{s}}^j(f)} \tag{6}$$

Here, we denote $|\mathcal{F}\{\boldsymbol{s}^j\}|$, the module of the Fourier transform of $\boldsymbol{s}^j$, as $\hat{\boldsymbol{s}}^j$. f_{HR} is the spectrum peak frequency which represents the subject heart rate defined in Eq. 7. r is the error toleration.

$$f_{HR} = \arg\max_f \mathcal{F}\{\boldsymbol{e}^j\} \tag{7}$$

To simplify the problem, we let y be a linear function, i.e.$y(g(.,.)) = \langle \boldsymbol{p}, g(.,.) \rangle$. Parameter $\boldsymbol{p} = [p_1, \ldots, p_N]$ could be regarded as the confidence vector which represents the patterns of signal strengths corresponding to N local face regions. Hence, the optimization problem can be written as follow

$$\arg\max_{\boldsymbol{p}} \sum_{j=1}^{J} \langle \boldsymbol{p}, g(\boldsymbol{s}^j, \boldsymbol{e}^j) \rangle \tag{8}$$

To normalize the confidence $\boldsymbol{p} = [p_1, \ldots, p_N]$ across all local face regions, we add a constraint to ensure that $\|\boldsymbol{p}\| \leq 1$.

In order to solve Eq. 8, we also need to obtain the "ground truth" $\boldsymbol{e}^j$ for the measurement of $g(\boldsymbol{s}^j, \boldsymbol{e}^j)$. Inspired by [20,30], we approximate $\boldsymbol{e}^j$ through PCA decomposition given signal $\boldsymbol{S} = [\boldsymbol{s}_1, \boldsymbol{s}_2, ...\boldsymbol{s}_N]^\intercal$ ($\boldsymbol{s}_i$ is centralized) and the corresponding confidence $\boldsymbol{p}$. Thus, the covariance matrix can be written as $\Sigma = \boldsymbol{S}^\intercal P \boldsymbol{S}$, where $P = diag(p_i^2)$. By applying standard PCA to Σ, we can reconstruct $\hat{\boldsymbol{E}} = [\hat{\boldsymbol{e}}_1, \ldots, \hat{\boldsymbol{e}}_N]^\intercal$ by $\hat{\boldsymbol{E}} = \boldsymbol{S}\Phi\Phi^\intercal$ where Φ is the eigenvectors correspond to the largest k eigenvalues that preserve α percent of the variance. Note that since $\boldsymbol{S}$ is constrained between a reasonable HR range in the rPPG extraction stage, $\boldsymbol{e}$ will also share the same property. Finally, we approximate $\boldsymbol{e}$ by

$$\boldsymbol{e} = \frac{1}{N} \sum_{i=1}^{N} \hat{\boldsymbol{e}}_i \tag{9}$$

Considering the estimation of $\boldsymbol{e}^j$ involves the inter-dependence between confidence $\boldsymbol{p}$ and signals $\boldsymbol{S}$, it may not be suitable to solve the objective function directly, with linear programming. Therefore, we apply an iterative approach, as summarized in Algorithm 1, to solve it by alternatively updating $\boldsymbol{p}$ and $\boldsymbol{e}$. At iteration t, we first update $\boldsymbol{e}^j(t)$ with the confidence $\boldsymbol{p}(t-1)$, and then update confidence $\boldsymbol{p}(t)$ with the updated "ground truth" $\boldsymbol{e}^j(t)$. When the convergence threshold δ is reached, we get the output confidence map $\boldsymbol{p}$.

Given the local rPPG confidence map $\boldsymbol{p}$, We could measure the confidence of $\boldsymbol{x}$ by computing each dimension's reliability. Following Eq. 3, we compute the confidence of $\boldsymbol{x}$ as

Algorithm 1. Local rPPG confidence learning

Input: Training signals $\boldsymbol{S} = [\boldsymbol{S}^1, \ldots, \boldsymbol{S}^J]$, converge threshold δ
Output: local rPPG confidence $\boldsymbol{p}$
$t = 1$, $\boldsymbol{p}(0) = \sqrt{N}/N$;
repeat
  for $j = 1$ *to* J **do**
    given $\boldsymbol{p}(t-1)$, apply PCA to $\Sigma = \boldsymbol{S}^{j\intercal} P \boldsymbol{S}^j$ where $P = diag(p_i^2(t-1))$;
    reconstruct $[\hat{\boldsymbol{e}}_1^j, \ldots, \hat{\boldsymbol{e}}_N^j]^\intercal = \boldsymbol{S}^j \boldsymbol{\Phi} \boldsymbol{\Phi}^\intercal$;
    update $\boldsymbol{e}^j(t)$ by computing Eq. 9;
  update $\boldsymbol{p}(t)$ by solving Eq. 8 given $[\boldsymbol{e}^1(t), \ldots, \boldsymbol{e}^J(t)]$;
until $|\boldsymbol{p}(t) - \boldsymbol{p}(t-1)| \leq \delta$;
return $\boldsymbol{p}(t)$;

$$\boldsymbol{q} = \bigcup_{\substack{i,j=1,\ldots,N \\ i \leqslant j}} p(s_i, s_j) \tag{10}$$

Here we assume the confidence of local regions are independent with each other, so, $p(s_i, s_j) = p_i p_j$.

Finally, we use SVM with RBF kernel for classification. In order to weaken the interference of corrupted local rPPG, we employ the joint confidence $\boldsymbol{q}$ to adjust the distance metric in RBF kernel as $RBF_{\boldsymbol{q}}(\boldsymbol{x}_i, \boldsymbol{x}_j) = e^{-\gamma D_{\boldsymbol{q}}(\boldsymbol{x}_i, \boldsymbol{x}_j)^2}$, where $D_{\boldsymbol{q}}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \sqrt{(\boldsymbol{x}_i - \boldsymbol{x}_j)^\intercal Q (\boldsymbol{x}_i - \boldsymbol{x}_j)}$ and $Q = diag(q_i)$.

5 Experiments

In this section, we first discuss the implementation details of the proposed method. After that, experiment datasets, testing protocol and baseline method will be introduced. Finally, we demonstrate and analyse the experiment results.

5.1 Implementation Details

Csiro face analysis SDK [29] is employed to detect and track 66 facial landmarks. In order to divide the face into local ROIs, 4 additional interest points are generated from the mid-point of landmarks (2, 33), (14 33), (1, 30) and (15, 30) [29]. As shown in Fig. 2, 22 unit ROIs are evenly defined as boxes. Finally, every 4 unit neighbor ROIs are combined to form 15 overlapped local ROIs (color boxes in Fig. 2). For rPPG extraction, we set the cutoff frequency as 40–180 beats/min through a bandpass filter. For local rPPG correlation model, we generate all the possible 120 ($C(N,2)+N = \frac{N!}{2!(N-2)!} + N$, $N = 15$) combinations from the 15 local rPPG signals and normalized them. For local rPPG confidence map, we set the error toleration $r = 3$ beats/min, convergence threshold $\delta = 10^{-3}$. In the estimation of $\boldsymbol{e}$, we set the $\alpha = 60\,\%$. Normally, eigenvectors that correspond to the largest 3 eigenvalues will be selected.

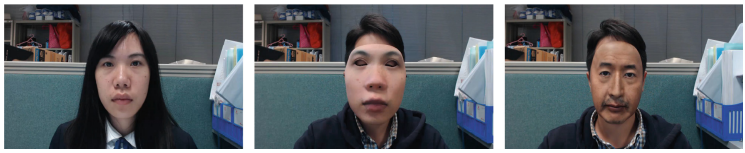


Fig. 3. Sample frames from supplementary dataset. The left image is the genuine face, the middle one is the Thatsmyface mask and the right one is the hyper-real mask from REAL-F (See Foonote 2)

5.2 Datasets

3DMAD. 3DMAD [13] is a public mask attack dataset built with the 3D masks from Thatsmyface.com. It contains 17 subjects, 3 sessions and total 255 videos (76500 frames). Each subject corresponds to 15 videos with 10 live faces and 5 masked faces. Videos are recorded through Kinect and contain color and depth information in 640*480 resolution. In our experiments, following [13], only the color information is used for comparison.

Supplementary Dataset. 3DMAD is a well organized dataset that contains large amount of videos from numerous of masks. But there are still some limitations: (1) Diversity of mask type is small. It only contains the masks from Thatsmyface.com. (2) All videos are recorded under the same camera setting through Kinect. To overcome these limitations, we create a supplementary (**SUP** for short) dataset to enlarge the diversity of mask types and camera settings. The SUP dataset contains 120 videos (36000 frames) recorded from 8 subjects. It includes 2 types of 3D masks: 6 from Thatsmyface.com and 2 from REAL-F[2]. Each subject has 10 genuine samples and 5 masked samples. All videos are recorded through Logeitech C920 web-camera in the resolution of 1280*720. Each video contains 300 frames and the frame rate is 25 fps. Image samples of the genuine video and 2 types of masked face videos are shown in Fig. 3. Noticed that 4 masks are not aligned with genuine subjects in the SUP dataset due to the budget issue. To our best knowledge, this adjustment will not affect the face anti-spoofing results since face anti-spoofing could be regarded as a 2-class classification problem without considering the subjects' identities. The SUP dataset will be public available.

By merging the supplementary dataset with the 3DMAD, the combined dataset (**COMB** for short) contains 25 subjects, 2 types of masks, and 2 camera settings, which has larger diversity that is close to the application scenario. Experiments are carried out on the COMB dataset and the SUP dataset.

5.3 Testing Protocols and Baseline Methods

Testing Protocol. We evaluate the effectiveness, and robustness of the proposed method under three protocols: (1) intra-dataset testing protocol, (2) cross-dataset testing protocol, (3) robustness evaluation.

[2] A super realistic 3D mask build from REAL-F: http://real-f.jp.

For intra-dataset testing protocol, we adopt leave-one-out cross validation (LOOCV) [13]. Different from [13], subjects in training set and development set are randomly[3] selected to avoid the possible affect of subjects sequence. For the combined dataset, we choose 8 subjects for training and 16 for development. For the SUP dataset, we randomly chose 3 subjects as training set and 4 as development set. To evaluate the influence of high quality masks from REAL-F, we test the performance by including and excluding the REAL-F masks in both datasets.

For the cross-dataset protocol, 3DMAD dataset and SUP dataset are involved. For the setting of training on 3DMAD and testing on SUP (**3DMAD to SUP** for short), we randomly select 8 subjects from 3DMAD as training set and use all subjects from SUP for testing. For training on SUP, testing on 3DMAD (**SUP to 3DMAD** for short), we randomly select 5 subjects from SUP as training set and use all in 3DMAD for testing.

In order to evaluate the robustness of the proposed method, we re-do the experiments under intra and cross testing protocols with different training scales. To avoid the resemblance affect of live faces and masks [13], we set the training data scale along subject units. For intra-dataset experiments on COMB dataset and SUP dataset, the training scales are set to 1 to 8 and 1 to 5, respectively. For the cross-dataset experiments of 3DMAD to SUP and SUP to 3DMAD, the training scale are set to 1 to 17 and 1 to 8, respectively.

False Fake Rate (FFR), False Liveness Rate (FLR), Half Total Error Rate (HTER) [13], ROC, AUC, and EER are employed for evaluation. For intra-dataset test, HTER is evaluated on testing set and training set. We name them as HTER_dev and HTER_test, respectively, for short.

Baseline Method. We select the Multi-Scale LBP [5] which achieves the best performance on 3DMAD 2D images [13] as the baseline. For a normalized face image, we extract $LBP^{u2}_{16,2}$, $LBP^{u2}_{8,2}$ from the entire image and $LBP^{u2}_{8,1}$ from the 3×3 overlapping regions. Therefore, one 59-bins, one 243-bins and nine 53-bins histograms feature are generated. We follow [13] on other setting details. Finally, histograms are concatenated as the final 833-dimensional feature representation.

5.4 Experimental Results

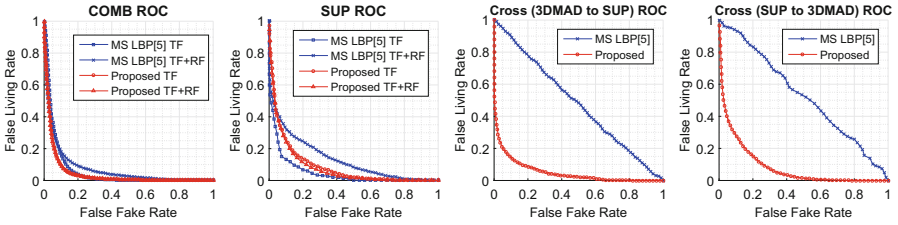
Intra-dataset results are given in Table 1, Fig. 4(a) and (b). We achieve the best performance on the combined dataset as well as the supplementary dataset, which justifies the effectiveness of the proposed method. Meanwhile, from Fig. 4(a) and (b), the proposed method achieves close results no matter with or without the hyper-real masks from REAL-F. This justifies our analysis in Sect. 3 that the rPPG-based solution is independent to the mask's appearance quality. Note that the MS-LBP drops (e.g., 8.4 % AUC on SUP and 1.3 % AUC on COMB) when including the hyper-real REAL-F masks in both datasets.

[3] Due to random selection of training data and development data, at least 20 round are tested and averaged for each experiment.

This may justify our analysis that the texture-based method may not be discriminative on masks with good appearance quality. As shown in Fig. 3, REAL-F masks have highly realistic appearance. The face structures of REAL-F are precisely corresponded. Skin texture is highly restored including the wrinkles, freckles and visible capillary vessels. Interestingly, comparing with 3DMAD, the proposed method shows lower performance on high resolution dataset: SUP. We hypothesize that this is due to the camera setting. In fact, SUP is recorded with dark background. In order to achieve appropriate global exposure, the camera automatically adjust the gain setting, and the actual exposure rate is not sufficient to extract clear rPPG signal.

Table 1. Experiment results on COMB and SUP under intra-dataset test protocol.

	Combined dataset				Supplementary dataset			
	HTER_dev (%)	HTER_test(%)	EER (%)	AUC (%)	HTER_dev (%)	HTER_test (%)	EER	AUC (%)
MS-LBP [5]	13.1 ± 6.3	13.8 ± 19.4	13.6	92.8	19.5 ± 11.1	23.0 ± 21.2	22.6	86.8
Proposed	$\mathbf{9.2 \pm 2.0}$	$\mathbf{9.7 \pm 12.6}$	**9.9**	**95.5**	$\mathbf{13.5 \pm 4.7}$	$\mathbf{14.7 \pm 10.9}$	**16.2**	**91.7**



(a) COMB dataset (b) SUP dataset (c) 3DMAD to SUP (d) SUP to 3DMAD

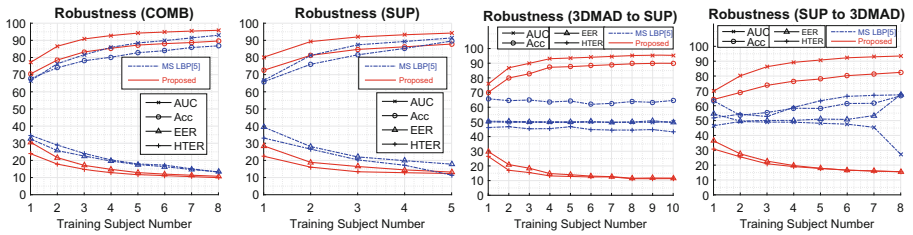
Fig. 4. ROC curves under intra-dabase and cross-dataset protocal. Note that the legend TF and RF means Thatsmyface mask and REAL-F mask.

Through the cross-dataset experiment results given in Table 2, Fig. 4(c) and (d), robustness of the proposed method have been demonstrated. This justifies the great adaptability of the proposed method when encountering different video qualities. Also, the dramatical performance decline of the MS-LBP may illustrate the analysis about over-fitting caused weak generalization ability. Note that training on 3DMAD achieves better performance than training on SUP. This may also because of the camera setting we discussed in intra-dataset results.

With the different training scale settings, the robustness of our proposed method has been illustrated. Figure 5 indicates that the proposed method could achieve good performance with small training data. With 5 subjects, the proposed method could nearly attain the best performance. It is because that, as analysed in Sect. 4, the local heartbeat pattern has small variance along different people and thereby is simple and easy to learn. This also justifies the feasibility of using rPPG as an intrinsic cue for face anti-spoofing.

Table 2. Experiment results between 3DMAD and SUP under cross-dataset test protocal.

	3DMAD to SUP			SUP to 3DMAD		
	HTER (%)	EER (%)	AUC (%)	HTER	EER	AUC (%)
MS-LBP [5]	46.5 ± 5.1	49.2	51.0	64.2 ± 16.7	51.6	47.3
Proposed	$\mathbf{11.9 \pm 2.7}$	**12.3**	**94.9**	$\mathbf{17.4 \pm 2.4}$	**17.7**	**91.2**



(a) COMB dataset (b) SUP dataset (c) 3DMAD to SUP (d) SUP to 3DMAD

Fig. 5. Robustness evaluation under intra-dabase and cross-dataset protocal

6 Conclusion and Discussion

In this paper, we propose to use rPPG as an intrinsic liveness cue for 3D mask face anti-spoofing. With the local rPPG correlation model and confidence measurement, the 3D mask can be detected effectively. Promising experimental results justify the feasibility of the proposed approach in combating 3D mask spoofing attack. Through cross-dataset experiment, the proposed method shows high potential on having a good generalization ability. The insights of this paper should have a substantial impact on the development of using rPPG as the liveness identifications for face anti-spoofing.

Besides, due to the expensive price of 3D mask, we only use 6 Thatsmyface masks and 2 REAL-F masks to increase the diversities of existing dataset. In future, more comprehensive analysis need to be evaluated with larger database which covers more interference and variation in application scenario, e.g., facial motion and light change.

Acknowledgement. We thank Baoyao Yang for her help on drawing Fig. 1. This project is partially supported by Hong Kong RGC General Research Fund HKBU 12201215, Academy of Finland and FiDiPro program of Tekes (project number: 1849/31/2015).

References

1. Rattani, A., Poh, N., Ross, A.: Analysis of user-specific score characteristics for spoof biometric attacks. In: 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 124–129. IEEE (2012)
2. Evans, N., Kinnunen, T., Yamagishi, J.: Spoofing and countermeasures for automatic speaker verification. In: INTERSPEECH, pp. 925–929 (2013)
3. Pavlidis, I., Symosek, P.: The imaging issue in an automatic face/disguise detection system. In: Proceedings of the IEEE Workshop on Computer Vision Beyond the Visible Spectrum: Methods and Applications, pp, 15–24. IEEE (2000)
4. Tan, X., Li, Y., Liu, J., Jiang, L.: Face liveness detection from a single image with sparse low rank bilinear discriminative model. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010. LNCS, vol. 6316, pp. 504–517. Springer, Heidelberg (2010). doi:10.1007/978-3-642-15567-3_37
5. Määttä, J., Hadid, A., Pietikainen, M.: Face spoofing detection from single images using micro-texture analysis. In: 2011 international joint conference on Biometrics (IJCB), pp. 1–7. IEEE (2011)
6. Anjos, A., Marcel, S.: Counter-measures to photo attacks in face recognition: a public database and a baseline. In: 2011 international joint conference on Biometrics (IJCB), pp. 1–7. IEEE (2011)
7. Zhang, Z., Yan, J., Liu, S., Lei, Z., Yi, D., Li, S.Z.: A face antispoofing database with diverse attacks. In: 2012 5th IAPR International Conference on Biometrics (ICB), pp. 26–31. IEEE (2012)
8. Pan, G., Sun, L., Wu, Z., Lao, S.: Eyeblink-based anti-spoofing in face recognition from a generic webcamera. In: IEEE 11th International Conference on Computer Vision, ICCV 2007, pp. 1–8. IEEE (2007)
9. de Freitas Pereira, T., Komulainen, J., Anjos, A., De Martino, J.M., Hadid, A., Pietikäinen, M., Marcel, S.: Face liveness detection using dynamic texture. EURASIP J. Image Video Process. **2014**(1), 1–15 (2014)
10. Kose, N., Dugelay, J.L.: Mask spoofing in face recognition and countermeasures. Image Vis. Comput. **32**(10), 779–789 (2014)
11. Yi, D., Lei, Z., Zhang, Z., Li, S.Z.: Face anti-spoofing: multi-spectral approach. In: Marcel, S., Nixon, M.S., Li, S.Z. (eds.) Handbook of Biometric Anti-Spoofing, pp. 83–102. Springer, London (2014)
12. Kose, N., Dugelay, J.L.: Shape and texture based countermeasure to protect face recognition systems against mask attacks. In: 2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 111–116. IEEE (2013)
13. Erdogmus, N., Marcel, S.: Spoofing face recognition with 3D masks. IEEE Trans. Inf. Forensics Secur. **9**(7), 1084–1097 (2014)
14. de Freitas Pereira, T., Anjos, A., De Martino, J.M., Marcel, S.: Can face anti-spoofing countermeasures work in a real world scenario? In: 2013 International Conference on Biometrics (ICB), pp. 1–8. IEEE (2013)
15. Wen, D., Han, H., Jain, A.K.: Face spoof detection with image distortion analysis. IEEE Trans. Inf. Forensics Secur. **10**(4), 746–761 (2015)
16. Chingovska, I., Anjos, A., Marcel, S.: On the effectiveness of local binary patterns in face anti-spoofing. In: 2012 BIOSIG-Proceedings of the International Conference of the Biometrics Special Interest Group (BIOSIG), pp. 1–7. IEEE (2012)
17. Verkruysse, W., Svaasand, L.O., Nelson, J.S.: Remote plethysmographic imaging using ambient light. Opt. Express **16**(26), 21434 (2008)

18. de Haan, G., Jeanne, V.: Robust pulse rate from chrominance-based rPPG. IEEE Trans. Bio-Med. Eng. **60**(10), 2878 (2013)
19. Li, X., Chen, J., Zhao, G., Pietikainen, M.: Remote heart rate measurement from face videos under realistic situations. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition, pp. 4264–4271, June 2014
20. Lempe, G., Zaunseder, S., Wirthgen, T., Zipser, S., Malberg, H.: ROI selection for remote photoplethysmography. In: Meinzer, H.-P., Deserno, T.M., Handels, H., Tolxdorff, T. (eds.) Bildverarbeitung für die Medizin 2013, pp. 99–103. Springer, Heidelberg (2013)
21. Menotti, D., Chiachia, G., Pinto, A., Schwartz, W.R., Pedrini, H., Falcao, A., Rocha, A.X.: Deep representations for iris, face, and fingerprint spoofing detection. IEEE Trans. Inf. Forensics Secur. **10**(4), 864–879 (2015)
22. Kollreider, K., Fronthaler, H., Faraj, M.I., Bigun, J.: Real-time face detection and motion analysis with application in liveness assessment. IEEE Trans. Inf. Forensics Secur. **2**(3), 548–558 (2007)
23. Wang, T., Yang, J., Lei, Z., Liao, S., Li, S.Z.: Face liveness detection using 3D structure recovered from a single camera. In: 2013 International Conference on Biometrics (ICB), pp. 1–6. IEEE (2013)
24. Komulainen, J., Hadid, A., Pietikainen, M.: Context based face anti-spoofing. In: 2013 IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS), pp. 1–8. IEEE (2013)
25. Poh, M.Z., McDuff, D.J., Picard, R.W.: Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. Opt. Express **18**(10), 10762–10774 (2010)
26. Lewandowska, M., Ruminski, J., Kocejko, T., Nowak, J.: Measuring pulse rate with a webcama non-contact method for evaluating cardiac activity. In: 2011 Federated Conference on Computer Science and Information Systems (FedCSIS), pp. 405–410. IEEE (2011)
27. Tulyakov, S., Alameda-Pineda, X., Ricci, E., Yin, L., Cohn, J.F., Sebe, N.: Self-adaptive matrix completion for heart rate estimation from face videos under realistic conditions. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2016
28. Kumar, M., Veeraraghavan, A., Sabharwal, A.: DistancePPG: robust non-contact vital signs monitoring using a camera. Biomed. Opt. Express **6**(5), 1565 (2015)
29. Cox, M., Nuevo-Chiquero, J., Saragih, J., Lucey, S.: CSIRO face analysis SDK, Brisbane, Australia (2013)
30. Wang, W., Stuijk, S., de Haan, G.: Exploiting spatial redundancy of image sensor for motion robust rPPG. IEEE Trans. Biomed. Eng. **62**(2), 415–425 (2015)