Chapter 13
Remote Blood Pulse Analysis for Face Presentation Attack Detection

Guillaume Heusch and Sébastien Marcel

Abstract In this chapter, the usage of Remote Photoplethysmography (rPPG) as a mean for face presentation attack detection is investigated. Remote photoplethysmography consists in retrieving the heart-rate of a subject from a video sequence containing some skin, and recorded at a distance. To get a pulse signal, such methods take advantage of subtle color variation on skin pixels due to the blood flowing through vessels. Since the inferred pulse signal gives information on the liveness of the recorded subject, it can be used for biometric presentation attack detection (PAD). Inspired by work made for speaker presentation attack detection, we propose to use long-term spectral statistical features of the pulse signal to discriminate real accesses from attack attempts. A thorough experimental evaluation, with different rPPG and classification algorithms is carried on four publicly available datasets containing a wide range of face presentation attacks. Obtained results suggest that the proposed features are effective for this task, and we empirically show that our approach performs better than state-of-the-art rPPG-based presentation attack detection algorithms.

13.1 Introduction

As face recognition systems are used for authentication purposes more and more, it is important to provide a mechanism to ensure that the biometric sample is genuine. Indeed, several studies showed that existing face recognition algorithms are not robust to simple spoofing attacks. Even simple display of a printed face photograph can fool biometric authentication systems. Nowadays, more sophisticated attacks could be performed by using high-quality silicone masks for instance [1]. Therefore, a remote authentication mechanism based on the face modality should take such

G. Heusch (✉) · S. Marcel
Idiap Research Institute, Rue Marconi 19, 1920 Martigny, Switzerland
e-mail: guillaume.heusch@idiap.ch

S. Marcel
e-mail: sebastien.marcel@idiap.ch

S. Marcel et al. (eds.), *Handbook of Biometric Anti-Spoofing*,
Advances in Computer Vision and Pattern Recognition,
https://doi.org/10.1007/978-3-319-92627-8_13

threats into account and provide a way to detect presentation attacks. In the last years, several methods to detect such attacks have been proposed, and are surveyed in both [2, 3]. Existing approaches can be roughly divided into two categories. The first category focuses on assessment of the liveliness of the presented biometric sample, by detecting blinking eyes [4] or exploiting motion information [5] for instance. The second category is concerned with finding the differences between images captured from real accesses and images coming from an attack. Representatives examples in this category include texture analysis [6], the usage of image quality measures [7] and frequency analysis [8]. However, current face presentation attacks methods suffers from their inability to generalize to different, or unknown attacks. Usually, existing approaches performs well on the same dataset they were trained on, but have difficulties when attack conditions are different [9]. However, a recent trend consists in deriving robust features that show better generalization abilities: examples can be found in [10, 11]. In the same spirit, presentation attack detection (PAD) based on remote blood pulse measurement is worth investigating: it should theoretically handle different attacks conditions well, since features does not depend on the type of attacks, but rather on properties of *bonafide* attempts.

13.1.1 Remote Photoplethysmography

Photoplethysmography (PPG) consists in measuring the variation in volume inside a tissue, using a light source. Since the heart pumps blood throughout the body, the volume of the arteria is changing with time. When a tissue is illuminated, the proportion of transmitted and reflected light varies accordingly, and the heart rate could thus be inferred from these variations. The aim of remote Photoplethysmography (rPPG) is to measure the same variations, but using ambient light instead of structured light and widely available sensors such as a simple webcam.

It has been empirically shown by Verkruysse et al. [12] that recorded skin colors (and especially the green channel) from a camera sensor contain subtle changes correlated to the variation in blood volumes. In their work, they considered the sequence of average color values in a manually defined region-of-interest (ROI) on the subject's forehead. After having filtered the obtained signals, they graphically showed that the green color signal main frequency corresponds to the heart rate of the subject.

Since then, there have been many attempts to infer the heart rate from video sequences containing skin pixels. Notable examples include the work by Poh et al. [13], where the authors proposed a technique where the color signals are processed by means of blind source separation (ICA), in order to isolate the component corresponding to the heart rate. In a similar trend, Lewandowska et al. [14] applied Principal Component Analysis (PCA) to the color signals and then manually selected the principal component that contains the variation due to blood flow. These two early studies empirically showed that the heart rate could be retrieved from video sequences of faces, but also highlight important limitations: the subject should be motionless, and proper lighting conditions must be ensured during the capture.

According to a recent survey [15], research in remote heart rate measurement has considerably increased in the last few years, most of which focuses on robustness to subject motion and illumination conditions. Since a large number of approaches have been proposed recently, they will not be discussed here. We refer the interested reader to [15, 16] for a comprehensive survey of existing algorithms. Current challenges in rPPG consists mainly of finding methods robust to a wide range of variability. For instance, de Haan et al. specifically devised a method to cope with subject motion in a fitness setting [17]. Also, it has been noted in [16] that different skin color tone affect the retrieved pulse signal. Lin et al. study the effect of different illumination conditions in [18]. Besides, video compression has also been identified as a limitations to retrieve reliable pulse signals [19].

13.1.2 rPPG and Face Presentation Attack Detection

Remote photoplethysmography is still an active research area, and that may explain that it has not been widely used in the context of face presentation attack detection yet. Moreover, and as noted in the previous section, main challenges to be addressed in this field (i.e. subject motion, illumination conditions and video quality) are usually present in a face recognition framework.

Despite its aforementioned limitations, rPPG has some potential for face presentation attack detection, as evidenced by previous work [20–22]. In this work, we thus propose to study pulse-based frequency features, as retrieved by rPPG algorithms, as a mean to discriminate real biometric accesses from presentation attacks. Indeed, in a legitimate, *bonafide* attempt, a consistent pulse signal should be detected, whereas such a signal should mostly consists of noise in case of a presentation attack. Furthermore, such approaches may have the desirable property to detect a wide range of attacks, since they do not rely on attack-specific information. They have the potential to overcome current limitations of classical PAD systems—relying on image quality or texture—through their better generalization abilities. Moreover, they are convenient, since they do not require user cooperation in assessing its liveness (challenge-response) nor do they necessitate additional hardware, such as devices studied in [23].

The typical workflow of a rPPG-based face presentation attack detection system is depicted in Fig. 13.1. Although several aspects of the whole system are considered in this work, our main contribution lies in the usage of long-term statistical spectral features, inspired by a recent work on speaker presentation attack detection [24]. Since these features are not specifically tailored to speech signals and are quite generic, we propose to use them on a pulse signal in the context of face presentation attack detection. Additionally, different rPPG algorithms as well as different classification scheme are studied. Extensive experiments are performed on four publicly available PAD databases following strict evaluation protocols. Besides, all the code needed to

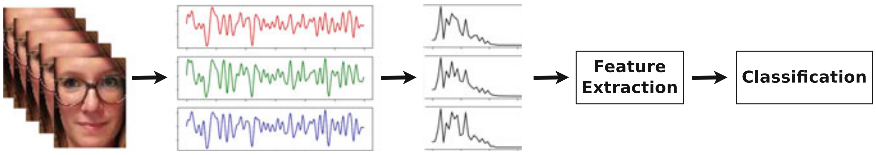


Fig. 13.1 Overview of a typical rPPG-based PAD system

reproduce presented results is made open-source and freely available to the research community.[1]

The rest of the paper is organized as follows: the next section presents prior work on remote physiological measurements for presentation attack detection. Then, proposed features are described, and considered rPPG algorithms as well as classification schemes are outlined. Databases and performances measures are presented in Sect. 13.4, before describing experiments and discussing obtained results. Finally, a conclusion is drawn and suggestions for future research are made in the last section.

13.2 Pulse-Based Approaches to Face Presentation Attack Detection

Remote Photoplethysmography has already been used in applications loosely related to face anti-spoofing. Gibert et al. [25] proposed a face detection algorithm, which builds a map of positive pulsatile response over an image sequence to detect the face. They even state that "Counterfeiting attempts using latex masks or images would be deceived if this map was taken into account". More recent work [26, 27] showed that detecting living skin using rPPG is feasible, at least in lab settings. However, using rPPG in the context of face PAD is still an emerging research area, as evidenced by the few number of previous works. At the time of writing, and to the best of our knowledge, only three studies using rPPG as a mean to detect presentation attack have been published. These previous relevant works are detailed below.

13.2.1 Liu et al.

Liu et al. [20] developed an algorithm based on local rPPG signals and their correlation. First, local pulse signals are extracted from different areas of the face. Usage of local signals is motivated for several reasons: first, it helps with robustness to acquisition conditions (illumination and subject's motion). Second, it can handle the case of a partially masked face, and finally, the strength of local rPPG signals are different

[1] Source code and results https://gitlab.idiap.ch/bob/bob.hobpad2.chapter13.

depending on the face area, but the strength pattern is the same across individuals. Local rPPG signals are extracted using the CHROM algorithm [28]. After having modeling the correlation of local pulse signals, a confidence map is learned and used for subsequent classification. Classification is done by feeding a Support Vector Machine (SVM) with local correlation models as features, and with an adapted RBF kernel using the confidence map as the metric. Their approach is evaluated on databases containing masks attacks only, namely 3DMAD [29] augmented with a supplementary dataset comprising six similar masks, plus two additional high-quality silicone masks. Obtained results on these different datasets, including cross dataset tests, show a good performance and hence validate the usage of pulse-based features to reliably detect masks presentation attacks. Unfortunately, the proposed algorithm is not assessed on traditionally used PAD databases, containing photo and video replay attacks.

13.2.2 Li et al.

Li et al. [21] suggest a relatively simple method to detect attacks using pulse-based features. First the pulse signal is retrieved using a simplified version of the algorithm presented in [30]. Three pulse signals—one for each color channel—are extracted by first considering the mean color value of pixels in a specific face area tracked along the sequence. Then, these colors signals are processed with three different temporal filters to finally get pulse signals, one in each color channel. Simple features are then extracted from each frequency spectra, and are concatenated before being fed to a linear SVM classifier. Experiments are again performed on 3DMAD, and also using the supplementary masks. Reported results show a better performance than [20], but do not seem to be directly comparable, since different experimental protocols were applied (training subjects were randomly chosen). An interesting point of this paper is that authors also report results on the MSU-MFSD database [7], and show that their method has difficulty to properly discriminate *bonafide* examples from video replay attacks.

13.2.3 PPGSecure

Nowara et al. [22] follow the same line of work as in [21], but considers the whole frequency spectrum derived from the intensity changes in the green color channel only. As in [20], this approach takes advantage of signals derived from different face areas, but also incorporates information from background areas (to achieve robustness to illumination fluctuations along the sequence). The final feature vector representing a video sequence is formed by concatenating the frequency spectra of pulse signals coming from five areas, three on the face (both cheeks and forehead) plus two on the background. Classification is then done either with a SVM or a random forest classifier. Experiments are performed on the widely used Replay-Attack database [6], but unfortunately, associated protocols have not been followed. Instead, the authors used a leave-one-subject-out cross validation scheme, which greatly increases the ratio of training to test data. Within this experimental framework, a perfect performance (i.e., 100%) accuracy is reported for both photographs and video attacks.

13.2.4 Discussion and Motivation

Although relevant, previous studies discussed here make it hard to objectively assess the effectiveness of rPPG-based approaches for face presentation attack detection. Indeed, performance is either reported on non-publicly available data or with different experimental protocols. As a consequence, it is difficult to compare published results with current state-of-the-art that relies on other means to detect attacks. A notable exception is [21], where authors reported results on the MSU-MFSD dataset. It also showed the limitation of such approaches, as compared to traditional face PAD approaches such as texture analysis.

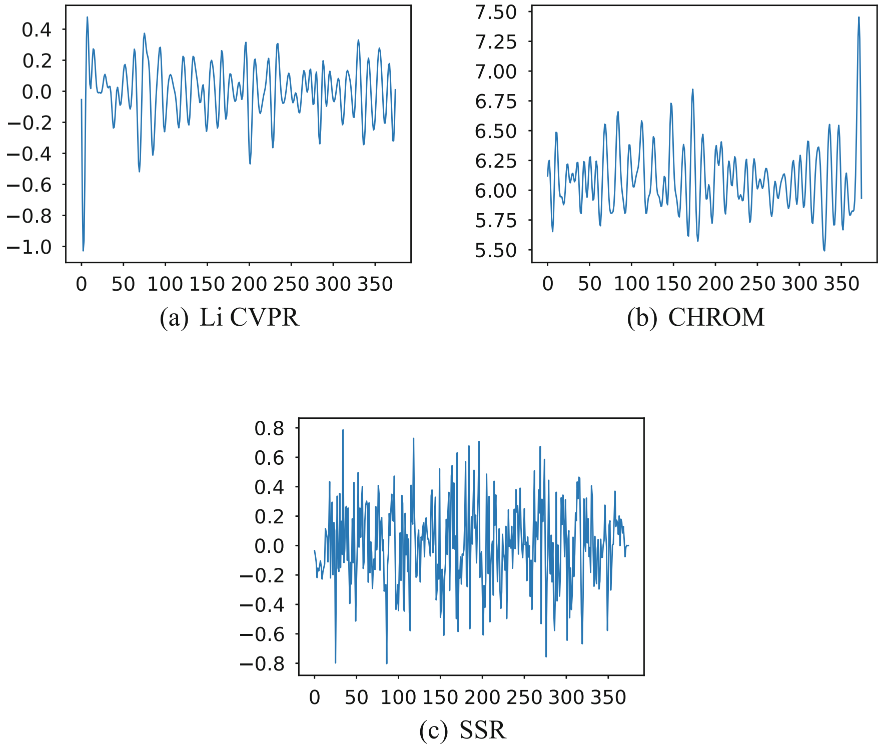
In this work, we hope to help foster research in this area by adopting a reproducible research approach. All the data and the software to reproduce presented results are available to the research community, easing further development in this field. Moreover, our proposed approach is assessed on four publicly available datasets, containing a wide variety of attacks (print and video replays of different quality, and mask attacks). The software package also comprise our own implementation of two other similar approaches, [21, 22], to which our proposed approach is compared.

13.3 Proposed Approach

In this contribution, we suggest to use Long-term spectral statistics (LTSS) [24] as features for face presentation attack detection. This idea was first developed in the context of speaker PAD, and managed to successfully discriminate real speakers from recordings in a speaker authentication task. The main advantage of such features is their ability to deal with any kind of signal and not necessarily speech.

Also, and since there exists a wide variety of rPPG algorithms, it seems important to consider more than one approach since they differ in the way the pulse signal is computed. This results in features that may be more suited to the task of presentation attack detection. To illustrate the difference, the retrieved pulse signals for a *bonafide* video sequence using the three investigated algorithms are shown in Fig. 13.2. One can clearly see that the pulse signals are not the same, depending on the used algorithm.

Furthermore, different classification algorithms are also investigated. In addition to classical two-class discriminative approaches, the usage of one-class classifiers



(a) Li CVPR

(b) CHROM

(c) SSR

Fig. 13.2 Pulse signals obtained with different rPPG algorithms

considering face presentation attack detection as an *outlier* detection problem are considered. Indeed, recent studies [31, 32] using this paradigm for face presentation attack detection showed promising results. Besides, one-class classifiers have been successfully applied for PAD on other modalities, such as speech [33] or fingerprint [34] and showed better generalization abilities. Furthermore, modeling real samples may be well-suited to pulse-based features, where properties of *bonafide* attempts only are considered.

13.3.1 Long-Term Spectral Statistics

In the context of pulse-based face PAD, and on the contrary to other approaches, prior knowledge on the characteristics of attacks is generally unknown. For instance, LBP-based systems intrinsically assume that texture of faces coming from presentation attacks are different that the one present in *bonafide* face images. These differences are manifold: this could be a lack of texture details on a mask for instance, or undesirable effects such as Moiré patterns or print artifacts in the case of replay and print attacks.

In our framework, the nature of the "pulse" signal extracted from an attack is unknown a priori. Therefore, no prior assumption on the negative class can be made: it is only assumed that signals differ in their statistical characteristics, irrespective of their content (i.e. we do not look specifically for periodicity for instance). As suggested in [24], the means and variances of the energy in different frequency bins provides such a generic characterization.

Long-term spectral statistics are derived by processing the original signal using overlapping temporal windows. In each window w, a N-point Discrete Fourier Transform is computed, and yields a vector $\mathbf{X}_w$ of dimension $k = 0, \ldots, N/2 - 1$ containing DFT coefficients. The statistics of frequency bins of the spectrum are considered using its log-magnitude. As in [24], whenever a DFT coefficient $|X_w(k)|$ is lower than 1, it is clipped to 1 such that the log-magnitude remains positive.

Using the set of DFT coefficient vectors $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_W$, the first and second order statistics of frequency components are computed as

$$\mu(k) = \frac{1}{W} \sum_{i=1}^{W} \log |X_w(k)| \tag{13.1}$$

$$\sigma^2(k) = \frac{1}{W} \sum_{i=1}^{W} (\log |X_w(k)| - \mu(k)) \tag{13.2}$$

for $k = 0, \ldots, N/2 - 1$. The mean and variance vectors are then concatenated to represent the spectral statistics of a given signal. As a result, the rPPG-based feature for classifying a video sequence consists of a single feature vector. The presentation attack detection is thus performed on the whole video sequence. In other approaches (i.e. texture or image quality-based), detection is generally peformed at the frame level. Long-term spectral statistics feature vectors are then used in conjunction with a classifier to reach a final decision on whether the given video sequence is a *bonafide* example, or an attack.

13.3.2 Investigated rPPG Algorithms

In this section, selected algorithms to retrieve a pulse signal are presented. Two of them, one proposed by Li et al. [30] and CHROM [28] already served as basis for face presentation attack detection in [21] and [20] respectively. The third one, Spatial Subspace Rotation (SSR) [35], has been chosen for both its original analysis (it does not rely on mean skin color processing but rather considers the whole set of skin color pixels) and its potential effectiveness, as demonstrated in [16].

Li CVPR

In this work, a simplified version of the rPPG algorithm originally developed in [30] has been implemented. This simplification has already been used for presentation attack detection in [21]. In particular, the correction for illumination and for motion are ignored. Basically, the pulse signal is obtained by first accumulating the mean skin color value across the lower region of a face in each frame and then to filter the color signal to get the pulse signal. In this work, instead of tracking the lower face region from frame to frame, it is computed at each frame by using a pre-trained facial landmark detector [36].

CHROM

The CHROM approach [28] is relatively simple but has been shown to perform well. The algorithm first finds skin-colored pixels in a given frame and computes the mean skin color. Then, the mean skin color value is projected onto a specific color subspace, which aims to reveal subtle color variations due to blood flow. The final pulse signal is obtained by first bandpass filtering temporal signals in the proposed chrominance colorspace, and then by combining these two filtered signals into one. Note that in our implementation, the skin color filter described in [37] has been used.

SSR

The Spatial Subspace Rotation (SSR) algorithm has been proposed in [35]. It considers the subspace of skin pixels in the RGB space and derives the pulse signal by analyzing the rotation angle of the skin color subspace in consecutive frames. To do so, the eigenvectors of the skin pixels correlation matrix are considered. More precisely, the angle between the principal eigenvector and the hyperplane defined by the two others is analyzed across a temporal window. As claimed by the authors, this algorithm is able to directly retrieve a reliable pulse signal, and hence no post-processing step (i.e., bandpass filtering) is required. Again, skin color pixels are detected using the filter proposed in [37].

13.3.3 Classification

Previous work in rPPG-based face presentation attack detection all rely on SVM—a classical discriminative algorithm—to perform classification of pulse-derived features. Although successful, we believe that choosing a suitable classifier should not be overlooked given the unpredictable nature of attacks. Therefore, a comparison of classification scheme is also performed. Since PAD is inherently a two-class problem, any binary classifier could potentially be used. The literature contains many examples and we refer the interested reader to [2, 3] for a comprehensive overview of existing approaches. In this work, three binary classification algorithms are applied to the proposed features: Support Vector Machine (SVM), Multi-Layer Perceptron (MLP) and Linear Discriminant Analysis (LDA). This choice of algorithms has been motivated by the fact that SVM seems to be the *defacto* standard in face PAD, and because it is used in all the previous work using pulse-based features. MLP and LDA have been chosen since they are used in conjunction with the proposed features in [24].

Although presentation detection attack is usually viewed as a two-class classification problem, it can also be seen as an *outlier* detection problem. According to [3], modeling the genuine examples distribution is a promising research direction, since one cannot anticipate every possible attack type. One-class classification has already been applied in the context of face presentation detection in [31, 32], where one-class SVM and Gaussian Mixture Models (GMM) have been used. These two algorithms are hence also applied to the proposed features here.

13.4 Experiments and Results

13.4.1 Databases

Replay-Attack

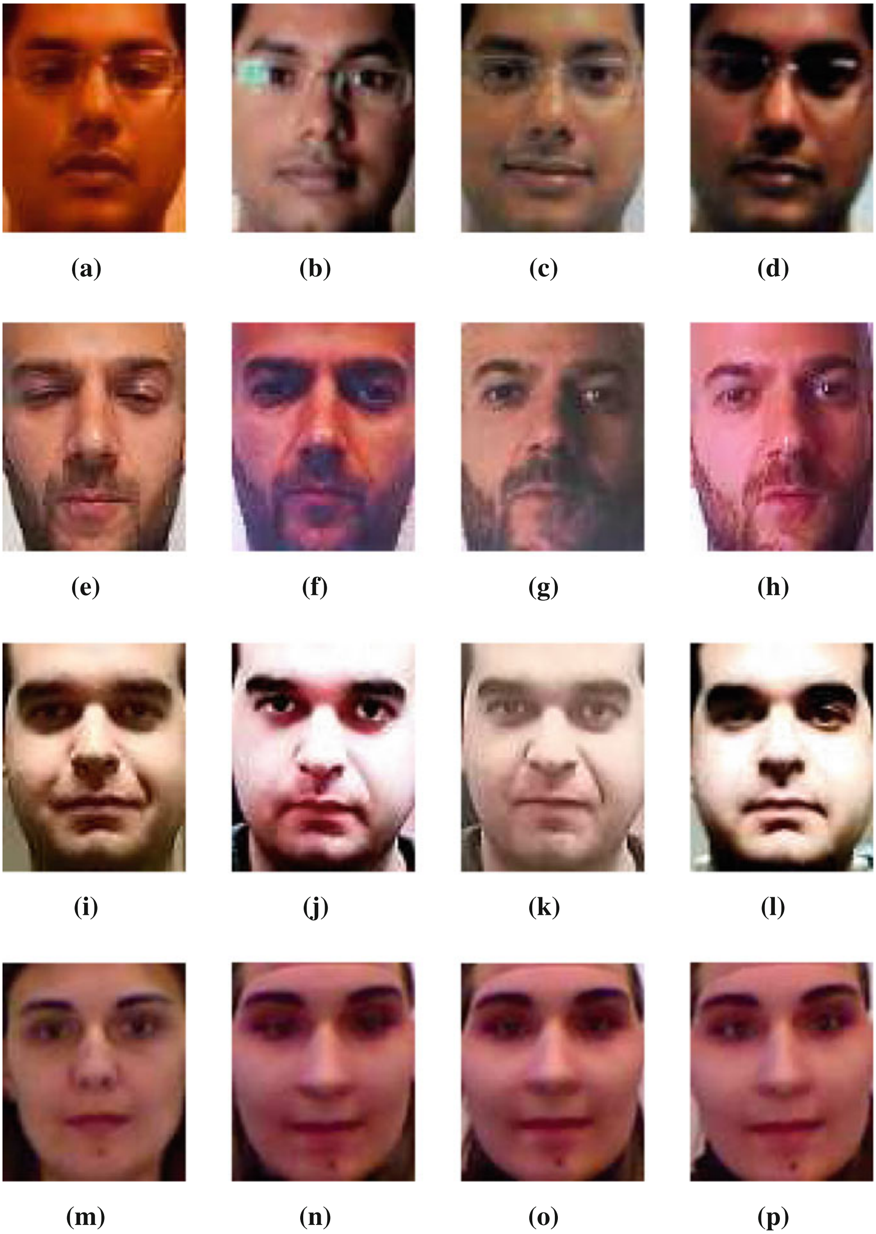
The Replay-Attack database was first presented in [6] and contains both *bonafide* attempts and presentation attacks for 50 different subjects. For each subject, two real accesses were recorded under different conditions, referred to as controlled and adverse. Presentation attacks were generated according to three different scenarios:

1. print: high-resolution photographs printed on A4 paper
2. mobile: photos and videos are displayed on an iPhone
3. highdef: photos and videos are displayed on an iPad

Also, two different conditions have been used to display attacks: either held by hand by an operator or attached to a fixed support in order to avoid motion. In total, there are 1200 video sequences, divided into training (360 seq.), development (360 seq.) and evaluation sets (480 seq.). The average sequence length is around 10 s (real accesses are longer and last about 15 s, whereas attacks last around 9 s). Although several protocols have been defined to assess the performance of face PAD algorithms, only the *grandtest* is used here, since it contains all the different attacks and hence allows to test various approaches for a wider range of attacks.

Replay-Mobile

The Replay-Mobile database [38] has been built in the same spirit as of the Replay-Attack database, but with higher quality devices to forge the different attacks. Indeed, attacks are here performed using either high-resolution videos presented on a matte screen or high quality photographs displayed on matte paper. This is done in order to minimize specular reflections, and hence to be closer to real access attempts. This dataset contains 1030 video sequences of 40 subjects, again divided into training (312 seq.), development (416 seq.) and evaluation (302 seq.) sets. The average length of the sequences is 11.8 s, and real accesses and attacks are usually of the same length. Experimental protocols have also been devised in a similar way than in Replay-Mobile, and again, we will restrict ourselves to the *grandtest* protocol, for the same reasons.



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

(e) (f) (g) (h)

(i) (j) (k) (l)

(m) (n) (o) (p)

Fig. 13.3 Examples of frames extracted from both *bonafide* accesses (first column) and presentation attacks (column 2–4). The first row shows examples from the Replay-attack database, the second one from replay-mobile, the third one from MSU-MFSD, and the fourth one from 3DMAD

MSU-MFSD

The MSU Mobile Face Spoofing Database has been introduced in [7]. It contains a total of 440 video sequences of 55 subjects, but only a subset comprising 35 subjects has been provided to the research community. Video sequences last around 9 s in average. This database also contains two types of attacks, namely high-quality photograph and video sequences. The publicly available subset specifies 15 subjects used for training and 20 subjects to perform evaluation: these specifications have not been followed here, since no development set is provided. Instead, we build a training set and a development set with 80 video sequences from 10 subjects each, and an evaluation set containing 120 sequences coming from the 15 remaining subjects.

3DMAD

The 3D Mask Attack Database (3DMAD) [29] is the first publicly available database for 3D face presentation detection. It consists in 15 videos sequences of 17 subjects, recorded thanks to a Microsoft Kinect sensor. Note that here, only color sequences are used. The sequences, which all last exactly 10 s, were collected in three different sessions: the first two are *bonafide* accesses and the third one contains the mask attack for each subject. The recordings have been made in controlled conditions and with uniform background. As in [29], we divided the database into training (105 seq. from 7 subjects), development and evaluation sets (75 seq. from 5 subjects in each). However, the random splitting has not been applied here: the training set simply contains the first seven clients, the development set is made with subjects 8–12, and the evaluation set with subjects 13–17. Examples of frames extracted from both real attempts and attacks for all databases can be found in Fig. 13.3).

13.4.2 Performance Measures

Any face presentation attack detection algorithm encounters two type of errors: either an attack is misclassified as a real access, or the other way around, i.e., *bonafide* attempts are wrongly classified as attacks. As a consequence, performance is usually assessed using two metrics. The Attack Presentation Classification Error Rate (APCER) is defined as the expected probability of a successful attack and is defined as follows:

$$APCER = \frac{\text{\# of accepted attacks}}{\text{\# of attacks}} \tag{13.3}$$

Conversely, the Bona Fide Presentation Classification Error Rate (BPCER) is defined as the expected probability that a *bonafide* access will be falsely declared as a presentation attack. The BPCER is computed as:

$$BPCER = \frac{\text{\# of rejected real accesses}}{\text{\# of real accesses}} \tag{13.4}$$

Note that according to the ISO/IEC 30107-3 standard, each attack type should be taken into account separately. We did not follow this standard here, since our goal is to assess the robustness for a wide range of attacks. Note also that these PAD specific measures relate to the more traditionally used False Acceptance Rate (equivalent to APCER) and False Rejection Rate (equivalent to BPCER).

To provide a single number for the performance, results are typically presented using the Half Total Error Rate (HTER), which is basically the mean of the APCER and the BPCER:

$$HTER(\tau) = \frac{(APCER(\tau) + BPCER(\tau))}{2} \quad [\%] \tag{13.5}$$

Note that the Half Total Error Rate depends on a threshold τ. Indeed, reducing the Attack Presentation Classification Error Rate will increase the Bonafide Presentation Classification Error Rate and vice-versa. The threshold τ is usually selected to minimize the Equal-Error Rate (EER, the operating point where APCER = BPCER) on the development set.

13.4.3 *Experimental Results*

In this section, the experimental framework and obtained results are presented. Implementation details are first discussed, before providing experimental results. In particular, a comparison of the proposed LTSS features is made with the spectral features proposed by both Li et al. [21] and Nowara et al. [22]. We then investigate the usage of different rPPG algorithms and classification schemes. Finally, an analysis of obtained results is made: it presents identified shortcomings and suggests directions for future research.

13.4.3.1 Implementation Details

For pulse retrieval, we used open-source implementation of selected rPPG algorithms[2] that have been compared for heart-rate retrieval in [39]. All algorithms have been used with their default parameters. Experiments have been performed on the four databases presented in Sect. 13.4.1, with their associated protocols. In particular, classifiers are trained using specified training sets, and various hyperparameters are optimized to minimize the EER on the development set. Finally, performance is assessed on the evaluation set. Experimental pipelines have been defined and performed using the bob toolbox [40, 41] and, as mentioned in Sect. 13.1, are reproducible by downloading the Python package associated with this contribution.

[2] https://pypi.python.org/pypi/bob.rppg.base.

Table 13.1 Performance of different features based on the frequency spectrum of the pulse signals. The HTER [%] is reported on the evaluation set of each databases

	Replay-Attack	Replay-Mobile	MSU-MFSD	3DMAD
Nowara et al. [22]	25.5	35.9	31.7	43.0
Li et al. [21]	27.3	30.7	23.3	29.0
Li CVPR + LTSS	**13.0**	**25.7**	**20.6**	**19.0**

13.4.3.2 Comparison with Existing Approaches

Here we present results for the proposed approach based on LTSS features and compare them with our own implementation of both previously published algorithms also using pulse frequency features [21, 22]. As features used in [21] come from pulse signals retrieved in three color channels, the only choice for the rPPG algorithm is Li CVPR [30]. The same approach has been made using the proposed LTSS features: they are computed from the frequency spectrum in each color channel and concatenated. Note that in the work of Nowara et al. [22], only the green channel has been considered.

For classification, a two-class SVM has been used to be consistent with previous studies. Therefore, the different systems mostly differs in the feature extraction step, making them easily comparable with each other. Table 13.1 shows the HTER performance of the different feature extraction approaches on the evaluation set of the different databases.

As can be seen, the proposed LTSS features achieve the best performance on all considered datasets, and provide a significant improvement over the similar investigated approaches. As compared to [21], where very simple statistics are used, long-term spectral statistics likely contain more information and are hence more suitable to reveal differences between pulse signals retrieved from real attempts and attacks. It also suggests that the temporal window-based analysis of frequency content is suitable: this is not surprising since pulse signals from real attempts should contain some periodicity, whereas pulse signals from attacks should not. Note finally that our implementation of Li's approach has a better performance on the MSU-MFSD dataset than the one reported in the original article [21]. Indeed, an EER of 20.0% is obtained, whereas authors reported an EER of 36.7% in [21].

When compared to features containing magnitude of the whole frequency spectrum in local areas [22], our proposed LTSS features performs consistently better, and by a large margin. This result is interesting for several reasons. First, features extracted from a single face region seem sufficient to retrieve valuable pulse information, as compared to features extracted from different local areas of the face. Second, embedding additional information (i.e features from the background) does not seem to help in this case. Finally, computing relevant statistics on the Fourier spectrum looks more suitable than using the whole spectrum as a feature.

Table 13.2 Performance when different algorithms are used to retrieve the pulse signal. The HTER [%] is reported on the evaluation set of each databases

	Replay-Attack	Replay-Mobile	MSU-MFSD	3DMAD
Li CVPR (green) + LTSS	16.1	**32.5**	**35.0**	17.0
CHROM + LTSS	20.9	38.1	50.6	29.0
SSR + LTSS	**5.9**	37.7	43.3	**13.0**

13.4.3.3 Comparison of Pulse Extraction Algorithms

In this section, we compare the different rPPG algorithms. Indeed, since they yield different pulse signals (see Fig. 13.2), it is interesting to see which one helps the most in discriminating *bonafide* attempts from presentation attacks. Since CHROM and SSR only retrieve a single pulse signal (and not three, as in [30]) LTSS features are derived from this single pulse signal only. For a fair comparison, and when using Li CVPR algorithm [30] for pulse extraction, only the pulse computed in the green channel is considered, since it has been shown that this color channel contains the most variation due to blood flow [16]. Table 13.2 reports the performance for different pulse extraction algorithms.

When comparing rPPG algorithms to retrieve the pulse signal, the SSR algorithm obtains the best performance on two out of four datasets. Actually, it has the overall best performance on both the Replay-Attack database with an HTER of 5.9% and on 3DMAD with an HTER of 13.0%. However, results on other databases do not show performance improvement as compared to the previous experiment, where LTSS features have been extracted and concatenated in three color channels. This suggests that in the context of PAD, all color channels carry valuable information.

13.4.3.4 Comparison of Classification Approaches

In this section, the different classifiers are compared. As mentioned in Sect. 13.3.3, several binary classification algorithms have been considered. The SVM is used with a classical RBF kernel in both two-class and one class settings. The MLP contains a single hidden layer and two outputs representing the two classes. Regarding Linear Discriminant Analysis, PCA is first applied to the features in order to ensure non-singularity of the within-class covariance matrix. Note also that in the LDA case features are projected to a single dimension, which is then directly used as a "score". Table 13.3 shows the obtained performance for different classification schemes.

It is clear from Table 13.3 that different classifiers obtain very different results on the same features. Within discriminant approaches, it is hard to define the most appropriate classifier for this task. They are quite close in terms of averaged performance over the four datasets: SVM has an average HTER of 19.8%, whereas MLP and LDA reach an average HTER of 21.8% and 22.4% respectively. Also, the optimal

Table 13.3 Performance of both tow-class and one-class classifiers. The HTER [%] is reported on the evaluation set of each databases

	Replay-Attack	Replay-Mobile	MSU-MFSD	3DMAD
SVM	**13.4**	26.0	20.6	19.0
MLP	27.8	27.7	**16.1**	**15.0**
LDA	24.6	**24.1**	23.3	17.0
GMM	21.6	54.1	35.6	44.0
OC-SVM	19.6	44.8	31.7	38.0

Table 13.4 HTER [%] performance on evaluation sets, with a breakdown on photo and video replay attacks

	Photo	Video
Replay-Attack	11.3	6.6
Replay-Mobile	19.0	26.5
MSU-MFSD	20.0	15.8

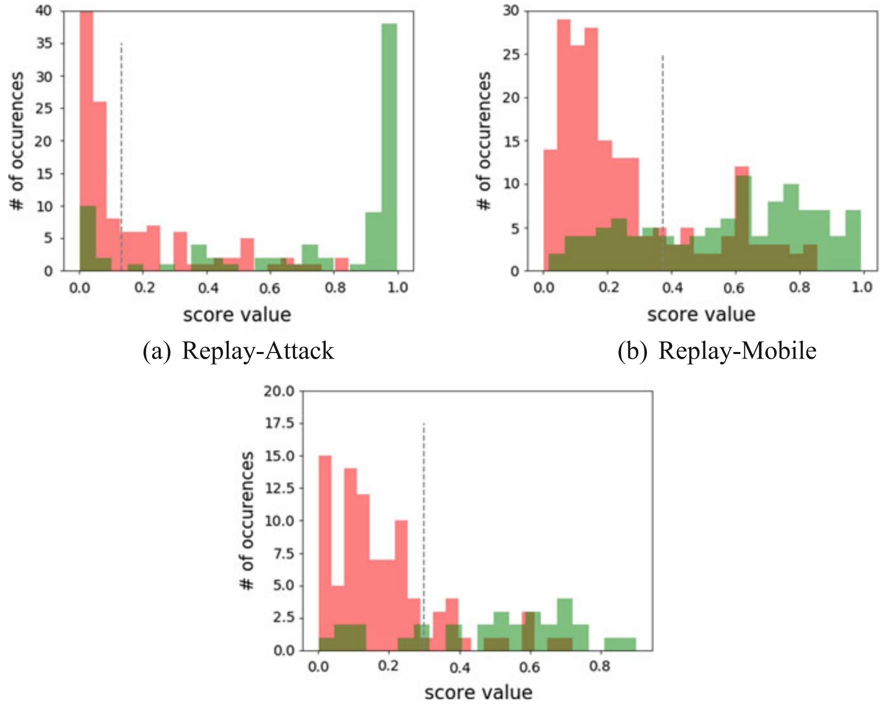
choice for the classifier is dependent on the database: this suggest that fusing different classifiers may be an interesting direction to investigate.

Table 13.3 also shows the poor performance obtained using the *outlier* detection approach. This may be explained by the lack of training data. Actually, modeling the distribution (GMM), or the support region (one-class SVM) of *bonafide* examples may be hard with few examples.

13.4.4 Discussion

In this section, a breakdown is made on the different attack types. This allows to better understand the behavior of our pulse-based face PAD approach, as well as to identify shortcomings, where future efforts should be made.

Table 13.4 shows the HTER of the proposed Li CVPR + LTSS system for two widely-used types of attack: photo and video replays. On the MSU-MFSD database, our approach performs better when dealing with video attacks, and this contradicts the result presented in [21]. Indeed, in the case of a photo attack, the image of the face is the same along the replayed sequence, therefore no pulse signal should be detected. Note that the same remark applies to the Replay-Attack database. This could maybe be explained by the motion introduced when the operator is holding the photograph in front of the camera, which may pollute the retrieved pulse signal. Also, some of the results reported on the Replay-Attack database in [22] exhibit the same trend: a better accuracy is sometimes observed on video attacks than on photo attacks.



(a) Replay-Attack

(b) Replay-Mobile

(c) MSU-MFSD

Fig. 13.4 Score values distribution of both *bonafide* accesses (green) and presentation attacks (red) on the evaluation set of the different databases. The dashed-line represents the decision threshold τ selected a priori on the development set. Note that for visualization purposes, the graph for Replay-Attack has been truncated. Actually the leftmost bin goes up to 300, meaning that most of the attacks have a very low score

Finally, the distribution of the scores obtained on the evaluation sets of the three databases containing both photo and video attacks are shown in Fig. 13.4 and provides two interesting insights:

1. Extracting reliable features from pulse signals is still a challenging problem for *bonafide* attempts. This is evidenced by the more uniform distribution of scores for genuine access (depicted in green in Fig. 13.4). This is especially true for both Replay-Mobile and MSU-MFSD databases. As a consequence, the BPCER is usually higher than the APCER.
2. On the other hand, proposed features are able to handle attacks pretty well: the distribution of attack scores (depicted in red in Fig. 13.4) spreads around a relatively low value on the left hand side of the histogram.

To further illustrate these observations, Fig. 13.5 shows example images, corresponding pulses and their respective frequency spectra for both *bonafide* examples

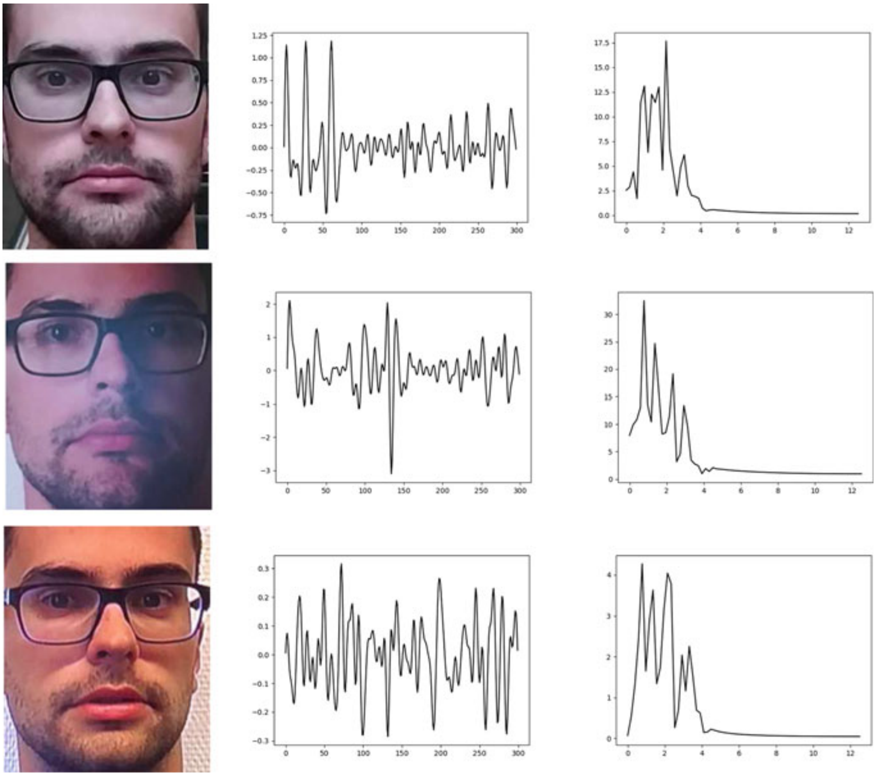


Fig. 13.5 Examples of images, retrieved pulses and their frequency spectrum for both real accesses and attacks from the Replay-Mobile database. The first row shows a legitimate access, the last two rows corresponds to a photo and a video attack respectively

(first row) and different presentation attacks (last two rows) of the Replay-Mobile database. One cannot clearly see differences in the frequency content between attacks and the real example. One would expect that for a real access, the corresponding rPPG signal would have a clear peak in the frequency spectrum that corresponds to the heart rate. In the example depicted in Fig. 13.5, it is actually the opposite: the pulse signal retrieved from the real access has more energy in high frequency components than the one in the photo attack. Note that high-frequency components are not present since the pulse signal is bandpassed; this may discard useful information to identify attacks, but recall that our goal is more oriented toward characterizing real accesses.

The same analysis has been made with mask attacks in the 3DMAD dataset and the score distribution is shown in Fig. 13.6. In this case, different observations can be made. Scores corresponding to *bonafide* examples are not that uniformly distributed and mainly lie on the right handside of the histogram, which is desirable. It means that for this dataset, extracted pulse-based features are more reliable than in previous case. This is not surprising, since sequences have been recorded under clean conditions

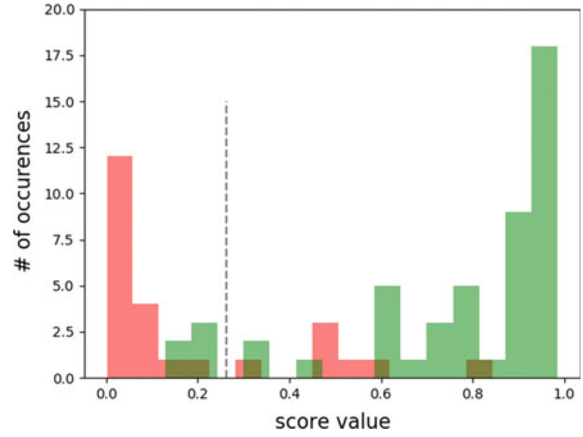


Fig. 13.6 Score values distribution of both *bonafide* accesses (green) and presentation attacks (red) on the evaluation set of the 3DMAD database. The dashed-line represents the decision threshold τ selected a priori on the development set

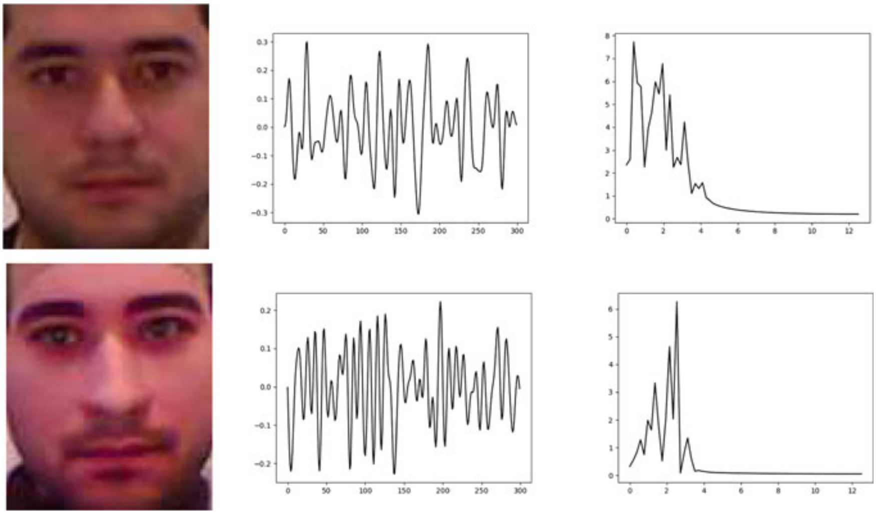


Fig. 13.7 Examples of images, retrieved pulses and their frequency spectrum for both real accesses and attacks from the 3DMAD database. The first row shows a legitimate access, and the second one is an attack

and do not contain as much variations as in other databases. Again, this suggest that illumination is an important factor for reliable pulse extraction.

Also, Fig. 13.7 shows example images, with their retrieved pulses and corresponding spectra for the 3DMAD database. Note here that the difference is easier to spot than in examples from the Replay-Mobile database (Fig. 13.5) and corresponds to expectations. In this case, one can clearly see the frequency component corresponding to the probable heart-rate of the subject (the leftmost peak of the spectrum) for the *bonafide* example. On the contrary, the signal retrieved from the attack is composed of higher frequencies, meaning that in this case, color variations should mainly be due to noise.

Although the proposed approach performs well as compared to other rPPG-based presentation attack detection, it does not reach state-of-the-art performance on these benchmarking datasets yet. Nevertheless, we believe that rPPG-based presentation attack detection systems have the potential to become successful, since there exists room for improvement.

First, and as evidenced in the previous analysis, a reliable pulse signal should be obtained. Current limitations of rPPG algorithms, and in particular illumination condition and compression have been identified and much effort is put on coping with this in current rPPG research. Second, existing approaches—including this one—consider relatively simple, hand-crafted features and progress can also be made here. For instance, Wang et al. successfully used more advanced spectral features in [27] to detect living skin. Moreover, recent advances in speaker presentation attack detection using convolutional neural networks (CNN) [42] show the superiority of such models over hand-crafted features. Finally, other classification approaches are to be studied yet. In particular, taking advantage of the temporal nature of the data using algorithms dedicated to time series, such as Hidden Markov Models or Recurrent Neural Networks, should be worth considering.

13.5 Conclusion

In this work, we studied the usage of remote photoplethysmography for face presentation attack detection. New features containing long term spectral statistics of pulse signals were proposed and successfully applied to this task. Experiments performed on four datasets containing a wide variety of attacks show that the proposed approach outperforms state-of-the-art pulse-based face PAD approaches by a large margin. Analysis of the results revealed that the greatest challenge for such systems is their ability to retrieve reliable pulse signals for *bonafide* attempts. This suggest that future work should first be directed towards improving rPPG algorithms in conditions suitable for PAD, where video quality is not necessarily sufficient for current approaches, and where both illumination variations and subject motion are present. Besides, there is also room for improvement in several other steps of the system. Automatically deriving pulse-based features, using convolutional neural networks for instance, and applying classification schemes tailored for time-series are, in our opinion, research directions worth investigating. Finally, such approaches have the potential to circumvent current limitations of face PAD systems. Actually, they may be well-suited to handle unknown attacks, since they only rely on properties exhibited in *bonafide* accesses, as opposed to approaches based on image quality or texture analysis.

Acknowledgements Part of this research is based upon work supported by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via IARPA R&D Contract No. 2017-17020200005. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the ODNI, IARPA, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon.

References

1. Liu S, Yang B, Yuen PC, Zhao G (2016) A 3D mask face anti-spoofing database with real world variations. In: IEEE conference computer vision and pattern recognition workshops (CVPRW), pp 1551–1557
2. Galbally J, Marcel S, Fierrez J (2014) Biometric antispoofing methods: a survey in face recognition. IEEE Access 2:1530–1552
3. Li L, Correia PL, Hadid A (2018) Face recognition under spoofing attacks: countermeasures and research directions. IET Biom 7(1):3–14
4. Pan G, Sun L, Wu Z, Lao S (2007) Eyeblink-based anti-spoofing in face recognition from a generic webcamera. In: International conference on computer vision (ICCV), pp 1–8
5. Anjos A, Marcel S (2011) Counter-measures to photo attacks in face recognition: a public database and a baseline. In: International joint conference on biometrics, pp 1–7
6. Chingovska I, Anjos A, Marcel S (2012) On the effectiveness of local binary patterns in face anti-spoofing. In: International conference of the biometrics special interest group, pp 1–7. IEEE
7. Wen D, Han H, Jain AK (2015) Face spoof detection with image distortion analysis. IEEE Trans Inf Forensics Secur 10(4):746–761
8. Caetano Garcia D, de Queiroz R (2015) Face-spoofing 2D-detection based on Moire-pattern analysis. IEEE Trans Inf Forensics Secur 10(4):778–786
9. de Freitas Pereira T, Anjos A, Martino JMD, Marcel S (2013) Can face anti-spoofing countermeasures work in a real world scenario? In: International conference on biometrics (ICB), pp 1–8
10. Patel K, Han H, Jain AK (2016) Cross-database face antispoofing with robust feature representation. In: Chinese conference on biometric recognition. Lecture Notes in Computer Science (LNCS), vol 9967. LNCS, pp 611–619
11. Patel K, Han H, Jain AK (2016) Secure face unlock: spoof detection on smartphones. IEEE Trans Inf Forensics Secur 11(10):2268–2283
12. Verkruysse W, Svaasand L, Nelson J (2008) Remote plethysmographic imaging using ambient light. Opt Express 16(26):21434–21445
13. Poh M, McDuff D, Picard R (2010) Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. Opt Express 18(10)
14. Lewandowska M, Ruminski J, Kocejko T, Nowak J (2011) Measuring pulse rate with a webcam—a non-contact method for evaluating cardiac activity. In: Proceedings federated conference on computer science and information systems, pp 405–410
15. McDuff D, Estepp J, Piasecki A, Blackford E (2015) A survey of remote optical photoplethysmographic imaging methods. In: IEEE international conference of the engineering in medicine and biology society (EMBC), pp 6398–6404
16. Wang W, den Brinker AC, Stuijk S, de Haan G (2017) Algorithmic principles of remote PPG. IEEE Trans Biomed Eng 64:1479–1491
17. de Haan G, van Leest A (2014) Improved motion robustness of remote-PPG by using the blood volume pulse signature. Physiol Meas 35(9):1913

18. Lin YC, Lin YH (2017) A study of color illumination effect on the SNR of rPPG signals. In: International conference on engineering in medicine and biology society (EMBC), pp 4301–4304
19. McDuff DJ, Blackford EB, Estepp JR (2017) The impact of video compression on remote cardiac pulse measurement using imaging photoplethysmography. In: IEEE international conference on automatic face and gesture recognition (AFGR), pp 63–70
20. Liu S, Yuen P, Zhang S, Zhao G (2016) 3D mask face anti-spoofing with remote photoplethysmography. In: European conference on computer vision (ECCV), pp 85–100
21. Li X, Komulainen J, Zhao G, Yuen PC, Pietikäinen M (2016) Generalized face anti-spoofing by detecting pulse from face videos. In: International conference on pattern recognition (ICPR), pp 4244–4249
22. Nowara EM, Sabharwal A, Veeraraghavan A (2017) PPGSecure: biometric presentation attack detection using photopletysmograms. In: IEEE international conference on automatic face and gesture recognition (AFGR), pp 56–62
23. Bhattacharjee S, Marcel S (2017) What you can't see can help you—extended-range imaging for 3D-mask presentation attack detection. In: International conference of the biometrics special interest group, pp 1–7
24. Muckenhirn H, Korshunov P, Magimai-Doss M, Marcel S (2017) Long-term spectral statistics for voice presentation attack detection. IEEE/ACM Trans Audio Speech Lang Process 25(11):2098–2111
25. Gibert G, D'Alessandro D, Lance F (2013) Face detection method based on photoplethysmography. In: IEEE international conference on advanced video and signal based surveillance, pp 449–453
26. Wang W, Stuijk S, de Haan G (2015) Unsupervised subject detection via remote PPG. IEEE Trans Biomed Eng 62(11):2629–2637
27. Wang W, Stuijk S, de Haan G (2017) Living-skin classification via remote-PPG. IEEE Trans Biomed Eng 64(12):2781–2792
28. de Haan G, Jeanne V (2013) Robust pulse rate from chrominance based rPPG. IEEE Trans Biomed Eng 60(10):2878–2886
29. Erdogmus N, Marcel S (2013) Spoofing in 2D face recognition with 3D masks and anti-spoofing with kinect. In: Proceedings of biometrics: theory, applications and systems (BTAS)
30. Li X, Chen J, Zhao G, Pietikainen M (2014) Remote heart rate measurement from face videos under realistic situations. In: IEEE Conference on computer vision and pattern recognition (CVPR)
31. Arashloo S, Kittler J (2017) An anomaly detection approach to face spoofing detection: a new formulation and evaluation protocol. IEEE Access, 80–89
32. Nikisins O, Mohammadi A, Anjos A, Marcel S (2018) On effectiveness of anomaly detection approaches against unseen presentation attacks in face anti-spoofing. In: International conference on biometrics (ICB)
33. Alegre F, Amehraye A, Evans N (2013) A one-class classification approach to generalised speaker verification spoofing countermeasures using local binary patterns. In: IEEE international conference on biometrics: theory, applications and systems (BTAS)
34. Ding Y, Ross A (2016) An ensemble of one-class SVMs for fingerprint spoof detection across different fabrication materials. IEEE international workshop on information forensics and security (WIFS)
35. Wang W, Stuijk S, de Haan G (2015) A novel algorithm for remote photoplethysmography: spatial subspace rotation. IEEE Trans Biomed Eng
36. King DE (2009) Dlib-ml: a machine learning toolkit. J Mach Learn Res 10:1755–1758
37. Taylor M, Morris T (2014) Adaptive skin segmentation via feature-based face detection. In: SPIE proceedings, real-time image and video processing, vol 9139
38. Costa-Pazo A, Bhattacharjee S, Vazquez-Fernandez E, Marcel S (2016) The replay-mobile face presentation-attack database. In: International conference of the biometrics special interest group

39. Heusch G, Anjos A, Marcel S (2017) A reproducible study on remote heart rate measurement. arXiv
40. Anjos A, Günther M, de Freitas Pereira T, Korshunov P, Mohammadi A, Marcel S (2017) Continuously reproducing toolchains in pattern recognition and machine learning experiments. In: International conference on machine learning (ICML)
41. Anjos A, El Shafey L, Wallace R, Günther M, McCool C, Marcel S (2012) Bob: a free signal processing and machine learning toolbox for researchers. In: ACM conference on multimedia systems (ACMMM)
42. Korshunov P, Gonçalves A, Violato R, Simões F, Marcel S (2018) On the use of convolutional neural networks for speech presentation attack detection. In: International conference on identity, security and behavior analysis (ISBA)