Pulse-based Features for Face Presentation Attack Detection

Guillaume Heusch and Sébastien Marcel
Idiap Research Institute
Rue Marconi 19, 1920 Martigny, Switzerland
{guillaume.heusch, sebastien.marcel}@idiap.ch

Abstract

In this contribution, we propose to tackle the face presentation attack detection (PAD) problem by using features derived from a pulse signal obtained through remote photoplesthymography (rPPG). Recent studies show that the pulse signal provides information on the liveness of a subject; hence it can be used to identify whether a recorded video sequence originates from a genuine user or is an attack. 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. Experiments are performed on different, publicly available databases and following associated protocols. 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.

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 spoofing attacks. Therefore, a remote authentication mechanism based on the face modality should take such 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 [10] and [13]. 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 [20] or exploiting motion information [3] for instance. The second category is concerned with finding the differences between images captured from real accesses and images coming from an attack. Representative examples in this category include texture analysis [5], the usage of image quality measures [26] and frequency analysis [4]. However, current presentation attack detection (PAD) 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 [21]. Therefore, PAD based on remote blood pulse measurement is worth investigating, since it should theoretically handle different attacks conditions well. Indeed, no assumptions are made on the nature of attacks. Rather, it relies on properties exhibited by *bonafide* attempts.

Photoplethysmography (PPG) measures 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 a pulse signal could thus be inferred from these variations. The aim of remote Photoplethysmography (rPPG) is to measure the same variations through a simple webcam. It has been empirically shown by Verkruysse *et al*. [23] that camera-recorded skin colors contain subtle changes correlated to the variation in blood volumes. Considering the sequence of average color values on the subject's forehead and filtering the obtained signals, they 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. According to a recent survey [17], the amount of work in remote heart rate measurement considerably increased in the last few years, focussing mostly on robustness to subject motion and illumination conditions. We refer the interested reader to [17] and [24] for a comprehensive survey of existing rPPG algorithms.

In this work, we propose to study pulse-based features, 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. As a consequence, such approaches have the potential to

978-1-5386-7180-1/18/$31.00

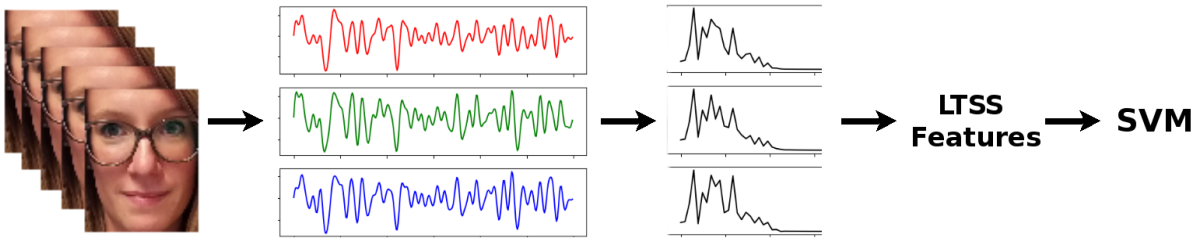


Figure 1. Overview of the proposed approach for Pulse-based Presentation Attack Detection.

detect a wide range of attacks, since they do not rely on attack-specific information such as texture. Our approach has been inspired by a recent work on speaker PAD [18], where long-term statistical spectral (LTSS) features are proposed. This approach show that first and second order statistics of the frequency spectrum of a speech signal are effective to detect presentation attack. Since these features are not specifically tailored to speech signals and are quite generic, we propose to apply the same approach, but on a pulse signal in the context of face PAD. The performance of our approach is assessed on four publicly available PAD databases following strict evaluation protocols. Besides, all the code needed to 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 face PAD. Then, the proposed approach is described, and considered rPPG algorithms are briefly outlined. Databases and performances measures are presented in Section 4. Experiments and results are discussed in Section 5. Finally, a conclusion is drawn and suggestions for future research are made in the last section.

2. Prior Work

At the time of writing, and to the best of our knowledge, only three studies using pulse-based features for face PAD have been published. Note that a very first attempt to use blood flow related information is briefly described in [6], but there is no further publications describing this approach. Previous works are described and briefly reviewed below.

Liu *et al*. [16] developed an algorithm based on local rPPG signals and their correlation. First, local rPPG signals are extracted using the CHROM algorithm [8] from different areas of the face. After having modeled 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 metric. Their approach is evaluated on databases containing masks attacks only, including 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.

Li *et al*. [15] 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 [14]. Three pulse signals - one for each color channel - are extracted by first considering the mean color value of pixels in a specific face area, that is tracked along the sequence. Then, these colors signals are processed with three different temporal filters to finally get pulse signals. 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 mask attacks. Reported results show a better performance than [16], but do not seem to be directly comparable, since different experimental protocols were applied. An interesting point of this paper is that authors also report results on the MSU-MFSD database [26], and show that their method has difficulty to properly discriminate *bonafide* examples from video presentation attacks.

Finally, Nowara *et al*. [19] consider the whole frequency spectrum derived from the intensity changes in the green color channel only. As in [16], this approach takes advantage of signals derived from different face areas, but also incorporates information from background areas (to be robust 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 5 areas, 3 on the face (both cheeks and forehead) plus 2 on the background. Classification is then again done with a SVM. Experiments are performed on the widely used Replay-Attack database [5], but unfortunately, associated protocols have not been followed. Instead, the authors used a leave-one-subject-out cross validation

[1] https://gitlab.idiap.ch/bob/bob.paper.btas_ivfib_2018

scheme, which greatly increases the ratio of training to test data. Within this experimental framework, 100% accuracy is reported for both photographs and video attacks.

These previous studies show that it is 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 [15], where authors reported results on the MSU-MFSD dataset and showed the limitation of such approaches. We hope to bridge this gap by presenting experiments on four publicly available datasets and by strictly following associated experimental protocols.

3. Proposed Approach

In this contribution, we suggest to use long-term spectral statistics (LTSS) [18]. This idea was first developed in the context of speaker presentation attack detection, 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 necesseraly speech.

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 $k = 0 \dots N/2 - 1$ containing DFT coefficients. The statistics of frequency bins of the spectrum are considered using its log-magnitude. As in [18], 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 $\mathbf{X}_1, \mathbf{X}_2, ...\mathbf{X}_W$, the first and second order statistics of frequency components are computed as:

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

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

for $k = 0 \dots N/2 - 1$. 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 classifiying a video sequence consists of a single feature vector, and the presentation attack detection is performed on the whole sequence and not on individual frames, as in other PAD approaches like image quality measures. Long-term spectral statistics feature vectors are then used in conjunction with a SVM to classify a given video sequence as a *bonafide* example or as an attack.

In this work, three different rPPG algorithms are considered to retrieve the pulse signal. Although their end goal is the same, they usually differ and yield different pulse signals, as can be seen in Figure 2. In the framework of PAD, such a comparison has never been done. Since the pulse signal is the first step of our proposed approach for PAD, we believe that different algorithms should be considered and compared.

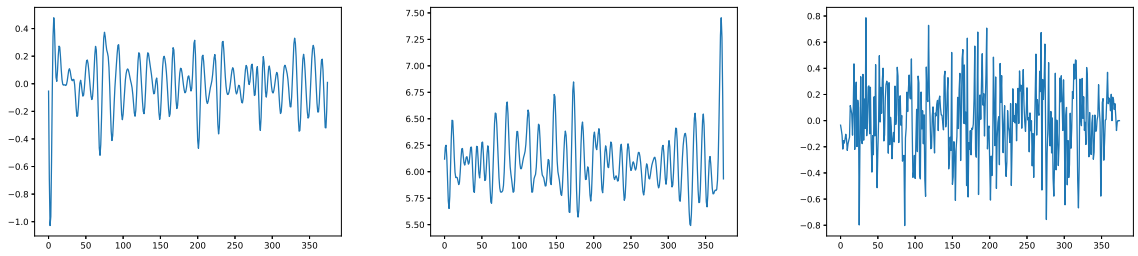
3.1. Investigated rPPG Algorithms

In this section, selected algorithms to retrieve a pulse signal are presented. Two of them, one proposed by Li *et al.* [14] and CHROM [8] already served as basis for face presentation attack detection in [15] and [16] respectively. The third one, Spatial Subspace Rotation (SSR) [25], 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 [24].

Li CVPR In this work, a simplified version of the rPPG algorithm originally developed in [14] has been implemented. This simplification has already been used for presentation attack detection in [15]. 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 [12].

CHROM The CHROM approach [8] 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 [22] has been used.

SSR The Spatial Subspace Rotation (SSR) algorithm has been proposed in [25]. 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



(a) Li CVPR (b) CHROM (c) SSR

Figure 2. Example of pulse signals retrieved from the same video sequence of a real attempt, with different rPPG algorithms.

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 [22].

4. Databases and Performance Measures

Replay-Attack The Replay-Attack database was first presented in [5] 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 different scenarios: high resolution photographs printed on A4 paper, plus photos and videos displayed on an iPhone or 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.). In this work, the *grandtest* experimental protocol is considered, since it contains all attacks.

Replay-Mobile The Replay-Mobile database [7] 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. Again, here we also consider the *grandtest* protocol.

MSU-MFSD The MSU Mobile Face Spoofing Database has been introduced in [26]. 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. 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 built 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) [9] 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. The sequences, which all last exactly 10 seconds, 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 [9], we divided the database into training (105 seq. from 7 subjects), development and evaluation sets (75 seq. from 5 subjects in each).

Performance Measures Any face presentation attack detection algorithm encounters two type errors: either *bonafide* attempts are wrongly classified as attacks, or the other way around, i.e. an attack is misclassified as a real access. 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 sucessful attack and is defined as follows:

$$APCER = \frac{\text{\# of accepted attacks}}{\text{\# of attacks}} \tag{3}$$

Conversely, the Bonafide Presentation Classification Error Rate (BPCER) is defined as the expected probability that

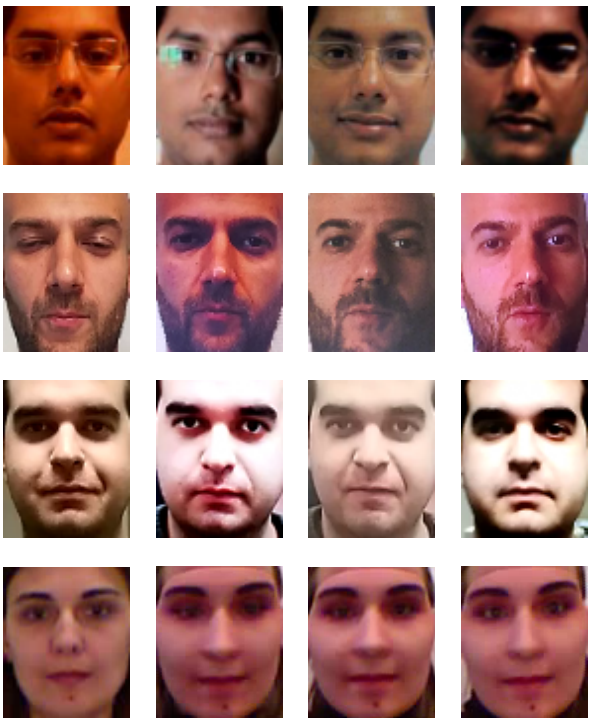


Figure 3. Examples of frames extracted from both *bonafide* accesses (first column) and presentation attacks (column 2 to 4). The first row shows examples from the Replay-Attack database, the second one from Replay-Mobile, the third one from MSU-MFSD, and the fourth one from 3DMAD.

a *bonafide* access will be falsely declared as a presentation attack. The BPCER is computed as:

$$BPCER = \frac{\text{\# of rejected real accesses}}{\text{\# of real accesses}} \quad (4)$$

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. 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:

$$HTER(\tau) = \frac{APCER(\tau) + BPCER(\tau)}{2} \quad [\%] \quad (5)$$

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

5. Experiments and 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. [15] and Nowara et al. [19]. Note that the approach proposed in [16] is not considered for comparison: it uses a correlation of local temporal signal as its main feature, whereas this work is more concerned with spectral features derived from pulse signals. We then investigate the usage of different rPPG algorithms. Finally, an analysis of obtained results is made, and presents identified shortcomings that should be addressed in future research.

5.1. Implementation Details

For pulse retrieval, we used an open-source implementation of selected rPPG algorithms [2] that have been compared for heart-rate retrieval in [11]. All algorithms have been used with their default parameters.

Experiments are performed on the four databases presented in Section 4, with their associated protocols. In particular, the classifier is trained using specified training sets, and 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 [2] [1] and, as mentioned in Section 1, are reproducible by downloading the Python package associated with this article.

5.2. Comparison of Spectral Features

Here we present results for the proposed approach based on LTSS features and compare them with our own implementation of algorithms proposed by Li *et al.* [15] and Nowara *et al.* [19]. As in [15], pulses are retrieved in each color channels using Li's CVPR rPPG method [14] and LTSS features derived from the three pulses are then concatenated. Note that in [19], only the green channel is considered. Table 1 shows the HTER performance on the evaluation set of the different databases. In following Tables, RA stands for Replay-Attack, RM for Replay-Mobile and MSU for MSU-MFSD datasets.

	RA	RM	MSU	3DMAD
Nowara et al. [19]	25.5	35.9	31.7	43.0
Li et al. [15]	27.3	30.7	23.3	29.0
Li CVPR + LTSS	**13.0**	**25.7**	**20.6**	**19.0**

Table 1. HTER [%] on the evaluation set of each databases.

As can be seen, the proposed LTSS features achieve the best performance on all datasets, and provide a large improvement over the similar investigated approaches. As compared to [15], where very simple statistics are used, it seems that long-term spectral statistics contain

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

more information and are hence more efficient at revealing differences between pulse signals retrieved from real attempts and attacks. It also suggests that the temporal window-based analysis of frequency content is suitable for pulse signals: this is not surprising since pulse signals from real attempts should contain some periodicity, whereas pulse signals from attacks should not. When compared to features containing magnitude of the whole frequency spectrum in local areas [19], our proposed LTSS features performs consistently better, 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. 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 [15]. Indeed, an EER of 20.0% is obtained, whereas authors reported an EER of 36.7% in [15].

5.3. Comparison of Pulse Extraction Algorithms

Here we compare the different rPPG algorithms. Indeed, since they yield different pulse signals (see Figure 2), it is interesting to see which one helps the most in discriminating *bonafide* attempts from presentation attacks. CHROM and SSR only retrieve a single pulse signal, therefore, LTSS features are derived from this single pulse signal as well. For a fair comparison, and when using Li CVPR algorithm [14] for pulse extraction, only the pulse computed in the green channel is considered. Table 2 reports the performance for different pulse extraction algorithms.

	RA	RM	MSU	3DMAD
Li CVPR + LTSS[3]	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**

Table 2. HTER [%] on the evaluation set of each databases.

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, more challenging 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.

[3]This result differs from Table 1 because LTSS are computed on the pulse signal *derived from the green channel only*.

5.4. Discussion

Time constraint Since the proposed approach relies on pulse signal analysis, a valid concern to be addressed is the required time needed to declare whether a transaction is a *bonafide* attempt or a presentation attack. Consequently, experiments were made with this constraint in mind. Pulse signals have been truncated before proceeding with LTSS feature extraction and classification. Note that the window size has been adjusted (if needed), such that the length of the window is at most one half of the signal's length. Figure 4 shows the performance of our approach as a function of elapsed time.

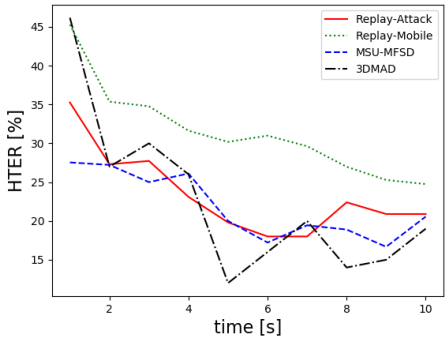


Figure 4. HTER as a function of elapsed time in seconds, for the different databases.

As expected, performance improves as time goes by, but not in a monotonic fashion. Except for the Replay-Mobile database, the performance, although fluctuating, reaches its optimum and remains quite stable after 4-5 seconds. Interestingly, a longer sequence does not necessarily mean an improved performance. This may be due to the introduction of more "noise" in *bonafide* attempts as the recording rolls on. Indeed, the recorded subject may be more prone to move, and illumination may slightly vary as well, posing difficulty in an accurate retrieval of the pulse signal.

Generic Considerations Finally, the distribution of the scores obtained on the evaluation set of the Replay-Mobile database is shown in Figure 5 and provides two interesting insights (similar observations have been made on other databases as well):

1. Extracting reliable features from pulse signals is still a challenging problem for *bonafide* attempts. This is evidenced by the almost uniform distribution of scores for genuine access (depicted in green in Figure 5).

2. On the other hand, proposed features are able to handle attacks pretty well: the distribution of attack scores (depicted in red in Figure 5) peaks at a relatively low value on the left hand side of the threshold.

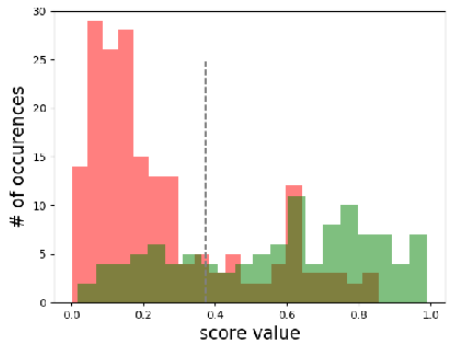


Figure 5. Score values distribution of both *bonafide* accesses (green) and presentation attacks (red) on the evaluation set of the Replay-Mobile database. The dashed-line represents the decision threshold τ selected *a priori* on the development set.

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. Nevertheless, we believe that rPPG-based presentation attack detection systems have the potential to become successful for this task. Such approaches have the advantage of handling unknown attacks, since they only rely on properties exhibited in *bonafide* accesses, as opposed to approaches based on image quality or texture analysis.

6. Conclusion

In this work, we studied the usage of rPPG 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, including a wide variety of attack, 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 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 automatically deriving pulse-based features, using convolutional neural networks for instance.

Acknowledgments

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] A. Anjos, L. El Shafey, R. Wallace, M. Günther, C. McCool, and S. Marcel. Bob: a free signal processing and machine learning toolbox for researchers. In *ACM Conf. on Multimedia Systems (ACMMM)*, Oct. 2012.

[2] A. Anjos, M. Günther, T. de Freitas Pereira, P. Korshunov, A. Mohammadi, and S. Marcel. Continuously Reproducing Toolchains in Pattern Recognition and Machine Learning Experiments. In *Intl Conf. on Machine Learning (ICML)*, Aug. 2017.

[3] A. Anjos and S. Marcel. Counter-Measures to Photo Attacks in Face Recognition: a Public Database and a Baseline. In *Intl Joint Conference on Biometrics*, pages 1–7, 2011.

[4] D. Caetano Garcia and R. de Queiroz. Face-Spoofing 2D-Detection Based on Moire-Pattern Analysis. *IEEE Trans. On Information Forensics and Security*, 10(4):778–786, 2015.

[5] I. Chingovska, A. Anjos, and S. Marcel. On the Effectiveness of Local Binary Patterns in Face Anti-spoofing. In *International Conference of the Biometrics Special Interest Group*, pages 1–7. IEEE, 2012.

[6] I. Chingovska, J. Yang, Z. Lei, D. Yi, S. Z.Li, O. Kähm, N. Damer, C. Glaser, A. Kuijper, A. Nouak, J. Komulainen, T. de Freitas Pereira, S. Gupta, S. Bansal, S. Khandelwal, A. Rai, T. Krishna, D. Goyal, M.-A. Waris, H. Zhang, I. Ahmad, S. Kiranyaz, M. Gabbouj, R. Tronci, M. Pili, N. Sirena, F. Roli, J. Galbally, J. Fierrez, A. Pinto, H. Pedrini, W. R. Schwartz, A. Rocha, A. Anjos, and S. Marcel. The 2nd Competition on Counter Measures to 2D Face Spoofing Attacks. In *Intl Conf. on Biometrics*, 2013.

[7] A. Costa-Pazo, S. Bhattacharjee, E. Vazquez-Fernandez, and S. Marcel. The Replay-Mobile Face Presentation-Attack Database. In *International Conference of the Biometrics Special Interest Group*, Sept. 2016.

[8] G. de Haan and V. Jeanne. Robust Pulse Rate From Chrominance Based rPPG. *IEEE Trans. On Biomedical Engineering*, 60(10):2878–2886, 2013.

[9] N. Erdogmus and S. Marcel. Spoofing in 2D Face Recognition with 3D Masks and Anti-Spoofing with Kinect. In *Biometrics: Theory, Applications and Systems (BTAS)*, 2013.

[10] J. Galbally, S. Marcel, and J. Fierrez. Biometric Antispoofing Methods: a Survey in Face Recognition. *IEEE Access*, 2:1530–1552, 2014.

[11] G. Heusch, A. Anjos, and S. Marcel. A Reproducible Study on Remote Heart Rate Measurement. 2017.

[12] D. E. King. Dlib-ml: a Machine Learning Toolkit. *Journal of Machine Learning Research*, 10:1755–1758, 2009.

[13] L. Li, P. L. Correia, and A. Hadid. Face Recognition Under Spoofing Attacks: Countermeasures and Research Directions. *IET Biometrics*, 7(1):3–14, 2018.

[14] X. Li, J. Chen, G. Zhao, and M. Pietikainen. Remote Heart Rate Measurement From Face Videos Under Realistic Situations. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2014.

[15] X. Li, J. Komulainen, G. Zhao, P.-C. Yuen, and M. Pietikäinen. Generalized Face Anti-spoofing by Detecting Pulse From Face Videos. In *Intl Conf. on Pattern Recognition (ICPR)*, pages 4244–4249, 2016.

[16] S. Liu, P. Yuen, S. Zhang, and G. Zhao. 3D Mask Face Anti-spoofing with Remote Photoplethysmography. In *European Conference on Computer Vision (ECCV)*, pages 85–100, 2016.

[17] D. McDuff, J. Estepp, A. Piasecki, and E. Blackford. A survey of remote optical photoplethysmographic imaging methods. In *IEEE Intl Conf. of the Engineering in Medicine and Biology Society (EMBC)*, pages 6398–6404, 2015.

[18] H. Muckenhirn, P. Korshunov, M. Magimai.-Doss, and S. Marcel. Long-term Spectral Statistics For Voice Presentation Attack Detection. *IEEE/ACM Transactions on Audio, Speech and Language Processing*, 25(11):2098–2111, Nov. 2017.

[19] E. M. Nowara, A. Sabharwal, and A. Veeraraghavan. PPGSecure: Biometric Presentation Attack Detection Using Photopletysmograms. In *IEEE Intl Conf on Automatic Face and Gesture Recognition (AFGR)*, pages 56–62, 2017.

[20] G. Pan, L. Sun, Z. Wu, and S. Lao. Eyeblink-based Anti-Spoofing in Face Recognition From a Generic Webcamera. In *Intl Conf. on Computer Vision (ICCV)*, pages 1–8, 2007.

[21] R. Ramachandra and C. Busch. Presentation Attack Detection Methods for Face Recognition Systems: A Comprehensive Survey. *ACM Computing Surveys*, 50(1):8:1–8:37, 2017.

[22] M. Taylor and T. Morris. Adaptive skin segmentation via feature-based face detection. In *SPIE Proceedings, Real-Time Image and Video Processing*, volume 9139, 2014.

[23] W. Verkruysse, L. Svaasand, and J. Nelson. Remote Plethysmographic Imaging Using Ambient Light. *Optics Express*, 16(26):21434–21445, 2008.

[24] W. Wang, A. C. den Brinker, S. Stuijk, and G. de Haan. Algorithmic Principles of Remote PPG. *IEEE Transactions on Biomedical Engineering*, 64:1479–1491, 2017.

[25] W. Wang, S. Stuijk, and G. de Haan. A Novel Algorithm for Remote Photoplethysmography: Spatial Subspace Rotation. *IEEE Transactions on Biomedical Engineering*, 2015.

[26] D. Wen, H. Han, and A. K. Jain. Face Spoof Detection with Image Distortion Analysis. *IEEE Trans. on Information Forensics and Security*, 10(4):746–761, 2015.