Continuous Presentation Attack Detection in Face Biometrics based on Heart Rate

Javier Hernandez-Ortega[1][0000−0001−6974−3900], Julian Fierrez[1][0000−0002−6343−5656], Ester Gonzalez-Sosa[1][0000−0002−2428−3792], and Aythami Morales[1][0000−0002−7268−4785]

Universidad Autonoma de Madrid, Madrid 28049, Spain
https://atvs.ii.uam.es/fierrez/

Abstract. In this paper we study face Presentation Attack Detection (PAD) against realistic 3D mask and high quality photo attacks in dynamic scenarios. We perform a comparison between a new pulse-based PAD approach based on a combination of a skin detector and a chrominance method, and the system used in our previous works (based on Blind Source Separation techniques, BSS). We also propose and study heuristical and statistical approaches for performing continuous PAD with low latency and false non-match rate. Results are reported using the 3D Mask Attack Database (3DMAD), and a self-collected dataset called BiDA Heart Rate Database (BiDA HR) including different video durations, resolutions, frame rates and attack artifacts. Several conclusions can be drawn from this work: 1) chrominance and BSS methods perform similarly under the controlled and favorable conditions found in 3DMAD and BiDA HR, 2) combining pulse information extracted from short-time sequences (e.g. 3 seconds) can be discriminant enough for performing the PAD task, 3) a high increase in PAD performance can be achieved with simple PAD score combination, and 4) the statistical method for continuous PAD outperforms the simple PAD score combination but it needs more data for building the statistical models.

Keywords: Face Presentation Attack Detection · Liveness Detection · Continuous Authentication.

1 Introduction

Nowadays, face is one of the most extended biometric traits along with iris and fingerprint. The causes of this spread are the inherent properties of face-based systems: samples can be acquired at a distance, passively, continuously, and using legacy hardware. Faces also contain highly discriminant features in order to achieve high accuracy rates when performing the recognition and verification tasks. Other significant reason of this spread is the deployment of biometrics broadly for the first time. Face-based systems are now present in numerous scenarios like medical applications, video-surveillance, mobile devices, e-commerce, etc.

Because of those reasons, attacks to face recognition systems are now more than ever, an important security issue. Among all the types of attacks, presentation attacks consist in showing an artifact to the sensor (e.g. a camera) for trying to disguise the attacker as a genuine user of the biometric system [10].

Presentation Attack Detection (PAD) techniques deal with these type of attacks. Even though high detection results can be obtained with these methods, the same PAD techniques may not be useful against all types of artifacts [14]. One of the most harsh menaces existing today are Mask Attacks, in which the presentation attack artifact is a 3D mask of a genuine user's face [6]. In these attacks, most PAD techniques successful against photo and video attacks, e.g. texture and depth based, become useless for high quality masks, because their similar properties (geometry, color, shape) to their real counterparts.

More recently, remote PhotoPlethysmoGraphy (rPPG) techniques [17], constisting in analysing videos for extracting the user's pulse signal, have been employed to analyze video sequences, proving to be an effective countermeasure against 3D mask attacks [12]. However, in order to achieve a robust estimation of the pulse signal, published approaches that use this method need long video sequences, good light conditions, are sensitive to failures in the face detection module, and also dependent to different acquisition sensors.

Current approaches like [18] perform a short-time approach to rPPG, more adequate to variable scenarios, in which the user or attacking conditions can change in the middle of the video sequence. In particular, in continuous scenarios where the attacker can enter at any time in a video stream, short-time approaches to PAD permit low latency PAD decisions. In this case holistic approaches are unable to give a continuous estimation of pulse and/or presentation attack probability, or PAD decisions with low latency. In addition, a short-time analysis of the rPPG signal also allows a better subsequent processing of the rPPG signals toward an overall more robust long-term estimation of the pulse.

Classic authentication schemes, in which users are authenticated employing an initial login stage, are able to stop unauthorized access attempts, but they are still unable of avoiding session "hijacking". In these attacks, a genuine user has been correctly authenticated and accepted by the PAD module, but after that, an attacker may be able to get control of his session. This problem is specially relevant in the field of mobile authentication, where the portability of the devices makes easier their theft or loss.

Continuous Authentication has emerged in biometrics to deal with the mentioned security problems in mobile devices and personal computers. These techniques consist in monitoring the user in a continuous way for verifying that the current user is the same who made the initial login, ideally in a transparent manner. For accomplishing this objective, biometrics such as the face [16] or the touch interaction [7] can be captured continuously without the user being aware. Our proposed approach for PAD follows the same continuous strategy, but in our case checking for PAD instead of identity. Please note that both identity and PAD can be incorporated in the loop, in a kind of continuous PAD and authentication scheme (see Fig. 1).

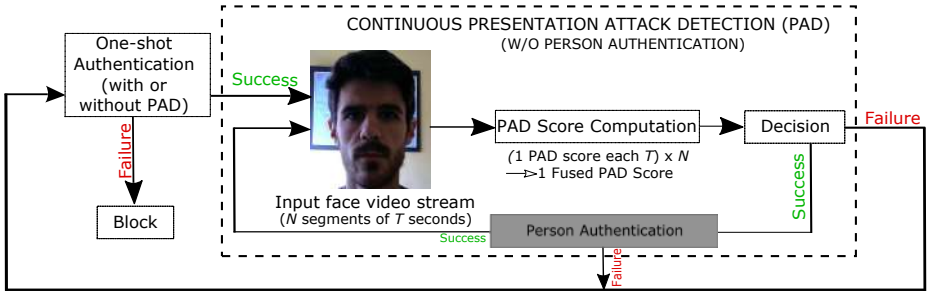

Fig. 1: **Proposed scheme for Continuous Presentation Attack Detection (PAD).** The authentication starts with a classic login (e.g. password, token or even a biometric trait), which may include or not PAD. If the login is successful, the continuous PAD loop starts working, generating PAD scores from the face video stream and deciding if a Presentation Attack has started or not. In the same loop we may also want to check also if the user is still the same (in gray), in a kind of Active Authentication scheme [16].

During a session hijacking, the attacker may be able to perform harmful actions, such as deleting or copying sensitive information, or installing a backdoor for granting future access to the compromised system. The latency of a continuous authentication method has the same level of criticalness than the accuracy rates, so a balance between usability and security must be achieved.

In this paper we: 1) present an algorithm based on rPPG for pulse detection applied to face Presentation Attack Detection (PAD); 2) study the performance of rPPG video-based continuous PAD, both in an existing benchmark (3DMAD) and a new dataset; and 3) test pulse-based continuous PAD in a scenario in which the attacking conditions vary over time.

The rest of this paper is organized as follows: Section 2 summarizes related works in rPPG and continuous authentication. Section 3 describes the proposed system. Section 4 describes the employed databases and the experimental protocol. Section 5 shows the results obtained. Finally, concluding remarks are drawn in Section 6.

2 Related Works

2.1 Remote Photoplethysmography

Photoplethysmography (PPG) [1] is a low-cost and noninvasive technique for measuring the cardiovascular Blood Volume Pulse (BVP) through variations in transmitted or reflected light. PPG can also be used to predict many vital health parameters such as blood pressure, heart rate (HR), hemoglobin and blood glucose level. Remote PPG (rPPG) consists in applying PPG techniques

to video sequences. These techniques look for changes in the color of the user's face that are caused by changes in the concentration of oxygen in the blood.

Related to our work, Poh et al. [17] measured HR from videos captured with a web-cam. They tracked the user's face and performed ICA to the RGB signal to separate the BVP chrominance signal from the other illumination variations and noise. On the other hand, the CHROM method [5] performs a linear combination of the spectrum bands to map the PPG signals to a space in which they are more robust to artifacts and noise.

Other works like [19] localize and track the information of certain facial regions instead of the entire face as there exist some zones that present higher variations in their color due to the pulsations. In [15] they use a special sensor that has the capability to capture other two aditional bands in the visible spectrum, since they have empirically proved to carry robustly the blood volume change information.

Regarding face PAD, when rPPG techniques are employed to estimate the pulse signal from a video sequence, the obtained result is highly different between the cases in which the recording contains a real face, and the cases with an attacking artifact (e.g. photos, videos, masks) [12].

Most research in this area employ self-collected datasets not publicly available. We decided to use 3DMAD as is one of the few public 3D mask PAD public datasets. It contains RGB videos of genuine users and of 3D mask attacks. We also employed a self-collected supplementary dataset in order to have larger RGB videos compared to the ones from 3DMAD. Larger recordings are necessary to measure the performance of continuous PAD techniques along time.

2.2 Continuous Authentication

A continuous authentication loop (see Fig. 1) can be added to any existing one-shot authentication system to improve its security. The most basic approach is based on using a single score in the authentication loop over time. The system generates a single score (i.e., $N = 1$ in Fig.1) each T seconds and decides if there is another user (or a presentation attack) based on that single authentication (or PAD) score.

The next level of complexity consist in combining several scores (i.e., $N > 1$) using different types of logic. The first approximation is based on calculating the arithmetic mean of several consecutive scores and taking a decision based on that combination. The combination can be done in a more complex way, for example considering that the confidence in the presence of the user decays when the time since the last authentication increases, with a function that decreases with time. On the other hand, not all the video frames have the same quality, for example due to occlusions, movement, blur, etc. Confidence functions can be built taking into account the quality level of the extracted signals [2]. However, these heuristic methods are very specific to each scenario, and do not have high generalization capabilities.

One can also use statistical methods for integrating multiple authentication or PAD scores in the continuous loop of Fig. 1, by using gallery information

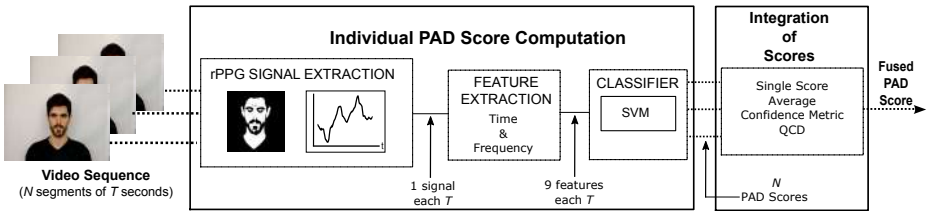

Fig. 2: **Architecture of the proposed module for continuous pulse-based face presentation attack detection (PAD Score Computation in Fig. 1)**. Given a facial video (N segments of T seconds each, with a time overlap α), the face is detected and rPPG-related features are extracted from the ROI in order to obtain an individual PAD score of each considered video segment(of T seconds). Then, the considered video segment generates an individual PAD score considering a database of real faces and mask attacks using a SVM. Finally the individual PAD scores are combined to derive the final fused PAD score corresponding to the full input Video Sequence.

to build models (typically real faces and attacks models for continuous PAD). Once the models are trained, the scores (single or multiple) that are extracted in the real environment are compared to the models in order to take a decision. A relevant work in this line is Quickest Change Detection (QCD). This technique has been employed successfully in mobile active authentication [16, 7] using multimodal data (i.e. face videos and touchscreen interaction).

3 Proposed Approach

The main purpose of the continuous PAD module proposed in Fig 1 consists in deciding if a video sequence contains images of real faces or images of presentation attacks. The architecture of this module is further detailed in Fig. 2. The first part is the rPPG signal extractor that obtains the pulse signal from the recordings. Once the rPPG signal is extracted from the video sequences, the second stage computes a set of features in order to distinguish between real faces and face attacks. The third step is a trained classifier that generates a score for each video sequence. The fourth and last stage integrates individual scores generated each T seconds, to form a final fused PAD Score each N individual scores.

3.1 rPPG Signal Extraction

The video sequences are generated from the input video stream by considering T seconds (with or without time overlap). The window length T allows to process larger or shorter pieces of the videos, having thus varying resolution in the final decision.

The rPPG signal generation is divided into two modules: skin detection and rPPG signal extraction:

Skin Detection. In the majority of the literature systems, the first stage consists in a face detection module (e.g. using the Viola-Jones algorithm) followed by ROI extraction. This stage selects one or several parts of the face that are assumed to contain robust information of the pulse signal. We applied this approach in our recent related work [11] and we have seen that it has several limitations such as: little robustness to movements, it can be difficult to implement, and it has a high computational load. Due to all these drawbacks, in this work we decided to apply the skin detector presented in [13] for getting our ROI. It transforms the video frames from the RGB color space to the YCrCb space. Their authors selected this color space as it has shown to have high discriminant properties for skin color modelling. The Y channel contains information of brightness while the Cr and Cb channels contain information about the differences between colors. A deeper description of the algorithm can be found in [13]. This algorithm skips a high number of pixels assuming that their values do not change within a small neighborhood. This approach reduce the CPU overload significantly making it suitable for real time video processing. Finally this algorithm does not depend of a face detection module, so it is more robust to user's movements.

rPPG Signal Extraction. Once the skin pixels have been located (see Fig. 3 (c) and (d) for examples), the next stage consists in extracting the rPPG signal from each considered segment (of T seconds). First, the raw values of the pulse signal are computed as the average intensity of the skin pixels. This calculation is made for each frame of the segment and for each the three color channels: Red, Green and Blue. The outputs are three rPPG sequences, one for each color channel. These raw rPPG signals contain not only the light variations produced by the blood volume changes, but also variations due to the external illumination and other noise sources. To reduce those undesired factors, in [11] we processed each channel as follows: a detrending filter for reducing the slow non-pulsating changes in the rPPG signal, a moving-average filter for eliminating random noise, and a band-pass filter for magnifying the frequency bands related to the usual pulse values. In this present work we decided to use the CHROM algorithm [5], which performs a linear combination of the three individual color channels into only one signal, robust to noise and external interferences [5]. This method also performs a frequency analysis of the signal for magnifying the bands related to a expected human pulse (between 0.6 Hz and 4 Hz).

3.2 Feature Extraction

In our previous work [11], used for reference, we decided to use the features from [12], where the authors transformed the signal from the spatial domain to the frequency domain using the FFT, and after that they estimated its Power

Table 1: Time and frequency features extracted from the postprocessed rPPG signal after applying the CHROM algorithm [5].

Domain	Feature	Description
Time	Zero Crossing Rate	Number of times the signal crossed the zero value
	Maximum/Minimum	Quotient between the temporal maximum and minimum
Frequency	P	Maximum power response
	R	Quotient of P and the total power in the 0.6 - 4 Hz frequency range
	Mean	Mean value of the signal
	Spectral Centroid	Mean value of each frequency component multiplied by its magnitude
	$\mathrm{N}_{\mathrm{max}}/N$	Sum of the N biggest values of the frequency signal divided by N
	LF Energy	Sum of the energy between 0 Hz and 4 Hz
	HF Energy	Sum of the energy between 2 Hz and 4 Hz

Spectral Density (PSD) distribution. Two features were extracted from each color band: the maximum power response P, and the ratio R between P and the total power in the 0.6-4 Hz frequency range.

For this work we decided to complement these two features P and R with other discriminant features that can give us more information about the rPPG signal in the time domain, following [4]. That work processed data from 3D accelerometer sensors, but their analysis is extrapolable to our rPPG signals. The final selected features can be seen in Table 1.

3.3 Classification

The last block of the presentation attack detection system is the classifier. Like in our reference work [11] we use Support Vector Machines (SVMs) as classifiers, in the present case considering the 9-dimensional features from Table 1 as input, and two classes as output: genuine face or face attack. Similar to related works [8], we use the signed distance to the separating surface obtained in the SVM training as output score of the Classifier in Fig. 2.

3.4 Integration of Individual Scores

In our experimental study we compare 4 different methods for the final stage in Fig. 2. The target of this stage is detecting the attacks as quick as possible (low Average Detection Delay, ADD), but trying to maintain a low value of real faces incorrectly detected as attacks (low False Non-Match Rate, FNMR). A deeper explanation of these terms can be found in [16].

Single Score. The first alternative only uses one input score for generating the fused PAD score (i.e., $N = 1$).

Mean Score. Individual PAD scores are averaged (applicable for $N > 1$).

Confidence-based Combination. A weighted sum of input scores is applied. The **first way** explored to define the weights consists in a time decay function. This function considers older samples as less reliable than the newer ones, since as time passes the conditions are high likely to have changed. The more recent scores will have a bigger weight. The **second way** is based in a rPPG quality measure [2]. In this work we decided to calculate a SNR value from each rPPG signal. In order to do that, we consider a perfect rPPG signal as sinusoidal, and all the other frequencies different to the one most relevant are considered as noise. The scores with a higher SNR will have a bigger weight when computing the sum.

Quickest Change Detection. QCD is a statistical method that first estimates match and non-match distributions of the scores, and then tries to detect the moment in which the new scores change from one distribution to the other. This type of approach needs prior data in order to build the match and non-match distributions. Some variants of QCD also require to know the probability of intrusion in advance, so we decided to implement the MiniMax QCD (MQCD) algorithm from [16], which only needs the score distributions.

4 Databases and Experimental Protocol

4.1 Databases

We use two different databases in order to compare results. The first is a public dataset named 3D Mask Attack Database (3DMAD) from the Idiap Research Institute [6]. We decided to use 3DMAD to enable direct comparison with related studies, primarily with our reference work [11]. The second database is a self-collected dataset named BiDA HR (BiDA Heart Rate database). It has been captured with the goal of complementing existing databases like 3DMAD, which have several limitations such as low resolution, few spectrum bands and short duration.

The 3D Mask Attack Database (3DMAD) [6] contains frontal-view recordings of 17 different users acquired using Microsoft Kinect. The dataset is composed by 3 different sessions, two with genuine accesses and one with 3D mask presentation attacks. Each session contains 5 videos of 10 seconds, captured at 30 frames per second, with a resolution of 640×480 pixels. The length of the videos is one important limitation of this database, as it would be desirable to have longer video sequences in order to study continuous authentication and continuous PAD methods.

The BiDA Heart Rate Database (BiDA HR) is a self collected dataset captured at the facilities of our research group at Universidad Autonoma de Madrid, in order to avoid the limitations from existent public databases. BiDA

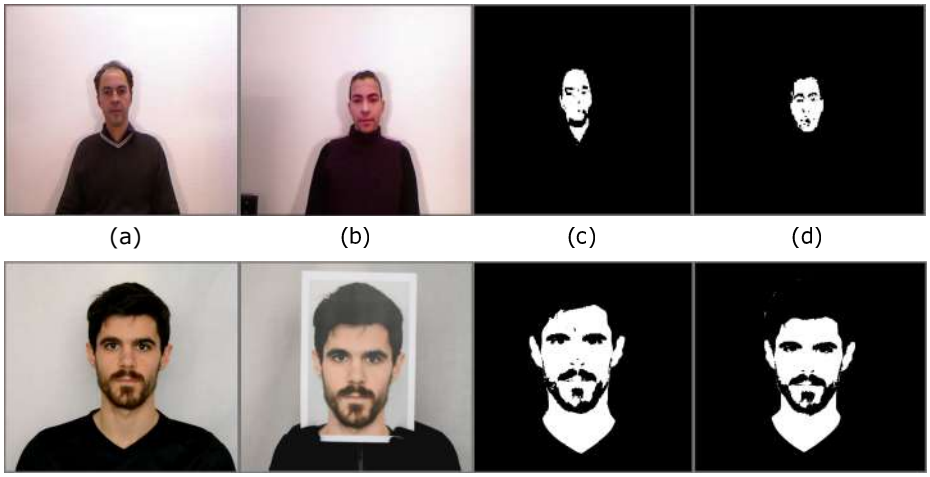

Fig. 3: **Datasets: 3DMAD (top) and BiDA HR (bottom)**. From left to right: (a) genuine access attempt, (b) presentation attack. We also show the outputs from the skin detection algorithm in (c) and (d) from a genuine access and a presentation attack respectively.

HR contains RGB, frontal-view, controlled, 60 second recordings of 10 different users, captured at 25 frames per second with 1920×1080 resolution (FullHD). It is a preliminary database and it has not been released yet. We are now capturing more samples to build a larger dataset. At its current state, the BiDA HR database is composed by 2 different sessions, one with real accesses and other with photo attacks. The artifacts of the attack attempts are HQ color printings of the faces (see Fig. 3). This way we are able to measure the performance of face PAD based on pulse detection with other type of easy-to-create spoofing artifacts different than 3D masks (the case in 3DMAD).

4.2 Experimental Protocol

From each rPPG signal we extracted the 9-dimensional feature vector described in Table 1. For classification in Fig. 2 we used Support Vector Machines with linear kernels and Cost parameter $C = 1000$ similarly to [11, 12].

Two experiments are conducted: first we emulate the results in our previous work [11]. This experiment does not try to show the performance of continuous authentication but it tries to compare the performances of both core rPPG algorithms. The second experiment consists in obtaining performance measures when using the proposed methods for continuous PAD presented in Sect. 3.4.

The experimental protocol is the same for both databases (3DMAD and BiDA HR). First of all, the whole dataset is divided into genuine samples and presentation attack samples. Then, for the first experiment, in order to train and

test the classifier, we use a Leave-One-Out Cross-Validation (LOOCV) protocol: for each subject in the database, we use all his feature vectors for testing a SVM model that has been trained with all the samples from the remaining users. The metric used to report results is the Equal Error Rate (EER in %). EER refers to the value where the Impostor Attack Presentation Match Rate (IAPMR, percentage of presentation attacks classified as real) and the False Non-Match Rate (FNMR, percentage of real faces classified as fake) are equal[1].

Results are obtained for several temporal window sizes: from 1 to 10 seconds in the case of 3DMAD, and also for 20, 40 and 60 seconds in the case of BiDA HR. For each temporal size T of the video segments, and considering a single video segment (i.e., $N = 1$ in Fig. 2), the EER has been calculated independently for all the subjects (each one of the LOOCV iterations). The individual results are then averaged to produce a single performance (mean and standard deviation of EER).

For the case of the continuous PAD experiments, we consider $N > 1$ in Fig. 2. In this case a PAD decision will be generated with a Delay of $D = N \times T$ seconds (video segments are not overlapped in time in our experiments).

Additionally, the QCD algorithm also needs prior data in order to build the match and non-match distributions. To compute those models, we use all data from 2 random users in each LOOCV iteration, who are left out of the LOOCV training and testing. In this case, additionally to the average EER rate, we have also computed an ADD-FNMR curve for varying temporal windows D. This curve is useful for showing the balance between the security and the usability of the continuous PAD approach proposed in Fig. 1.

Finally, for a deeper understanding of the QCD performance, we have also included some examples of the evolution of the fused PAD score during an example attack attempt. As the databases do not contain videos combining real faces and attacks, we have built videos concatenating a real access and an attack of the same user.

5 Results

5.1 Comparison with reference work

Table 2 shows the results of the comparison between the reference rPPG pipeline from [11] and the current work. Highlighted in bold are the best EER results for each value of the video length T. As can be seen in the table, none of the systems is absolutely better than the other in terms of performance. In general, the present system achieves lower EER rates than [11] when working with larger values of T (> 5 seconds), but the differences in the error rates are low. If the

[1] As error measures we have mentioned IAPMR and FNMR as defined and discussed by Galbally *et al.* [9]. Modifying the Decision Threshold until those error rates are equal we obtain the Presentation Attack Equal Error Rate, PAEER, defined and discussed in [9]. Here we follow [9] using PAEER to evaluate the presentation attacks, but calling it as EER for simplicity.

Table 2: **Comparison between the proposed rPPG face PAD and [11]** on 3DMAD and BiDA HR databases. The study has been performed changing the length T of the video sequences analyzed. Values in %. Lower values for each window length T are highlighted in bold.

3DMAD	Length T [s]	1	2	3	4	5	6	7	8	9	10
[11]	Mean EER [%]	**42.8**	45.0	37.8	**40.7**	**33.1**	29.7	**25**	26.1	24.1	22.1
	Std EER [%]	5.0	5.9	8.6	9.8	10.8	18.1	14.5	15.2	11.9	10.3
Present Work	Mean EER [%]	44.7	**42.2**	**37.3**	46.1	46.1	**28.8**	26.1	**25.8**	**22.3**	**18.8**
	Std EER [%]	4.1	6.7	8.5	5.9	5.45	11.8	13.1	12.2	12.3	13.4

BiDA HR	Length T [s]	1	2	5	10	20	30	40	50	60
[11]	Mean EER [%]	**46.9**	**45.7**	**42.1**	40.1	40.0	40.0	36.6	**30.0**	**25.0**
	Std EER [%]	3.9	5.1	9.5	9.6	14.0	21.1	20.5	25.8	26.3
Present work	Mean EER [%]	48.5	46.5	43.1	**38.6**	**38.9**	**31.2**	**30.8**	32.5	26.2
	Std EER [%]	2.4	2.7	6.3	11.3	10.1	15.6	18.4	22.9	23.1

databases contained less controlled conditions: more head motion, light changes, blur, etc, then we would expect more benefits from the skin detection and the CHROM algorithm proposed now, as they have shown to perform more robustly than [11] under these type of conditions.

Comparing the EER results obtained with 3DMAD data with those obtained with BiDA HR, there is a gap between performances, achieving lower rates in the case of 3DMAD. As we discussed in [11], this seems to be due to the lower frame rate of BiDA HR.

5.2 Continuous PAD

PAD Score Integration. Fig. 4 shows the PAD mean score combination from Sect. 3.4 on both databases. In that figure, the x axis corresponds to the values of $D = N \times T$, the delay for releasing the PAD decision (see Fig. 2), while the different curves represent the performances obtained with different temporal resolution T. It can be seen that, in general, the lowest EER (i.e., best PAD performance) is not obtained when using large T, but intermediate values (e.g. $T = 3$ s.). With even shorter values of T (1 s., or 2 s.) the amount of available scores within each decision window will be higher, but the reliability of each individual score will be lower (as can be seen in Table 2). On the other hand, the individual scores obtained with large values of T are the most reliable, but in this case there will be little data to combine within each decision window of size D.

The reader can notice that the EER obtained for specific D and T may be higher than the results showed in Table 2 for equivalent values (note that T in Table 2 should be compared to D in Fig. 4). For example, this is the case of T=1 and D=10 in Fig. 4 vs $T = 10$ in Table 2. While the classic approach

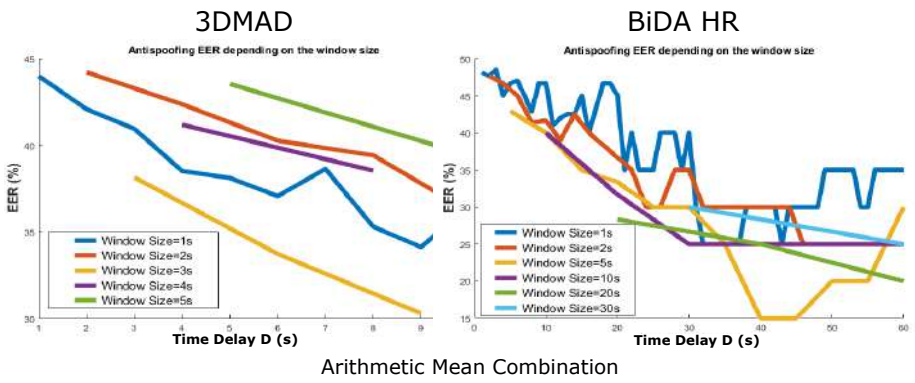

Fig. 4: **Mean score combination of individual PAD scores**, generated in temporal windows of varying duration on 3DMAD (left) and BiDA HR (right) databases. The x axis corresponds to the value of $D = N \times T$, the decision delay, while the different curves represent the performances obtained with different sizes of the temporal resolution T, and N is the number of individual PAD scores being combined.

(without the continuous loop in Fig. 1) is able to achieve a EER value of 18.8% at the 3DMAD database, the continuous PAD is only able of getting around 35%. However, the proposed continuous approach provides higher temporal resolution (decisions each second) and it is also able to improve the ongoing decisions by considering both old data and new one. The classic approach is only able to give one decision and only after the full 10 seconds have passed (high latency).

The best results from the 3 heuristic PAD methods compared (mean score, time based combination, and SNR-based combination) are obtained with the arithmetic mean. The SNR-based combination has failed to distinguish the samples with more quality from each recording. We think that modeling the pulse signal as a sinusoidal is not capturing appropriately the nature of a high quality pulse signal. With more accurate models it might be possible to achieve lower error rates using this approach. Finally, the temporal confidence fails to achieve lower EER rates than the other methods, and we think this is mainly caused by the limitations of the employed databases. This confidence measure is designed to deal with variable scenarios in which the conditions (attack/non attack) change within the same video. However these databases only contain recordings of attacks and real attempts performed separately. If this method was applied to a more realistic scenario, the results might be better than the ones obtained with the other two heuristic methods.

Quickest Change Detection (QCD). Fig. 5 shows the ADD-FNMR curves obtained with the QCD algorithm for the 3DMAD and BiDA HR databases. The different pairs of values of ADD and FNMR have been computed varying the

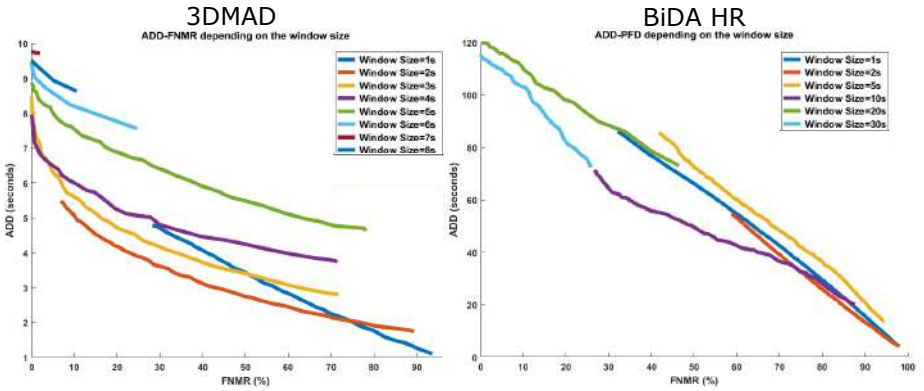

Fig. 5: **Average Presentation Attack Detection Delay (ADD) vs False Non-Match Rate (FNMR)** obtained on the 3DMAD (left) and BiDA HR (right) databases, for different temporal segments T.

decision threshold for each temporal segment T. The results from both databases show the same properties. In these curves, the best choice of ADD-FNMR depends of the real application of the system. Generally, a lower area under the curve is an indicator of a better performance. As can be seen in Fig. 4, the best results (as a balance of usability and security) are obtained with medium values of T, as it provides a good balance between the reliability of the scores and a low latency. When working with large values of T it is impossible to achieve low ADDs because of the inherent latency due to the analyzed temporal segments, of duration T. This limitation does not exist when working with small T, but these approaches are unable to obtain FNMR values as low as the obtained with a bigger T, due to the smaller reliability of the individual PAD scores being fused.

Finally, Fig. 6 shows an example of the evolution of the liveness scores obtained with the MQCD algorithm. In this scenario we wanted to simulate a real situation in which the attacker puts on the mask inside the video, so we have concatenated two different videos from the same user (a real access and an attack attempt). The higher the scores the higher the estimated probability of a presentation attack. Thanks to the MQCD and its low latency approach, the PAD score is able to evolve over the video, and can be compared to a threshold to detect the intrusion with low latency.

6 Conclusions and Future Work

In this paper, we have studied face Presentation Attack Detection (PAD) based on remote PhothoPlethysmoGraphy (rPPG) or, in other words, video-based heart rate estimation. We have extracted pulse information from facial videos from two different databases: 3DMAD and BiDA HR. These databases contain videos with different resolutions, frame rates, durations, and spoofing artifacts.

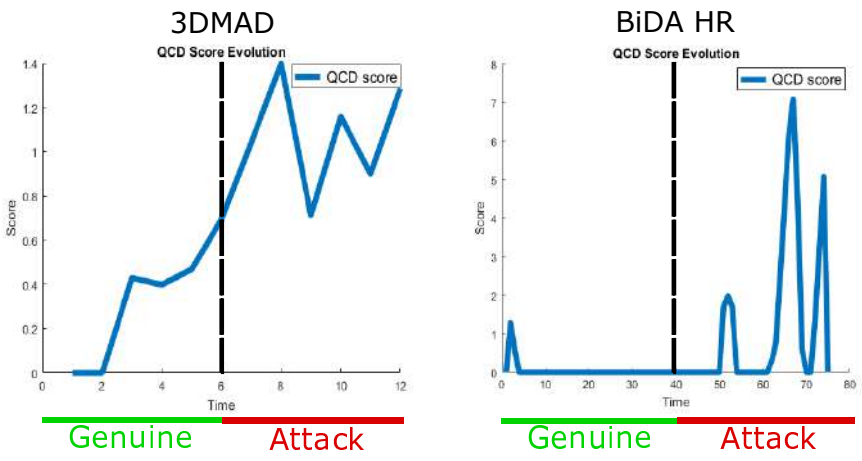

Fig. 6: **Temporal evolution of the fused PAD scores (QCD) in a variable attack scenario**. The attacker puts on the mask inside the example video. Results using data from 3DMAD (left) and BiDA HR (right) are shown.

We have compared the performance of a new rPPG system based on a combination of a skin detector and a chrominance method, and the system used in our previous work [11], which was based on Blind Source Separation techniques. Even though the chrominance-based system is more robust to variable light conditions, blur, and other factors, in this work both systems perform in a similar way due to the controlled conditions found in 3DMAD and BiDA HR.

We have also analyzed several approaches for low-latency continuous PAD. The first approach combines individual PAD scores with simple rules obtained from consecutive small video segments. The arithmetic mean of consecutive scores outperforms SNR-based and temporal-based score combination functions. The second approximation to continuous PAD uses a Quickest Change Detection algorithm (MQCD) for getting a balance between low attack detection delays (ADD) and low false positive rates (FNMR). Best results were obtained by generating individual PAD scores from video segments of around 3 seconds. We also discussed a possible time-variant attack scenario in which the attacker puts on the mask in the middle of the video. In this scenario, the advantages of a short-time rPPG analysis can be fully exploited.

Future work includes: 1) Improving the baseline system for getting lower EER with short videos (e.g. using video magnification techniques [3]). 2) Capturing a larger database with a higher number of users, more variate spoofing artifacts, and also more challenging conditions (like ambient illumination, blur, occlusions, etc). 3) Accomplishing a more in depth study of the performance when changing spatial and temporal resolution of videos. And 4) developing more robust quality metrics in rPPG [2] for score combination in continuous PAD and continuous authentication [8].

References

1. Allen, J.: Photoplethysmography and its application in clinical physiological measurement. Physiological measurement **28**(3) (2007)
2. Alonso-Fernandez, F., Fierrez, J., Ortega-Garcia, J.: Quality Measures in Biometric Systems. IEEE Security Privacy **10**(6), 52–62 (2012)
3. Bharadwaj, S., Dhamecha, T.I., Vatsa, M., Singh, R.: Computationally efficient face spoofing detection with motion magnification. In: IEEE Conf. on Computer Vision and Pattern Recognition Workshops (CVPRW). pp. 105–110 (2013)
4. Dargie, W.: Analysis of time and frequency domain features of accelerometer measurements. In: Int. Conf. on Compt. Comms. and Networks. IEEE (2009)
5. De Haan, G., Jeanne, V.: Robust pulse rate from chrominance-based rPPG. IEEE Trans. on Biomedical Engineering **60**(10), 2878–2886 (2013)
6. Erdogmus, N., Marcel, S.: Spoofing face recognition with 3D masks. IEEE Trans. on Information Forensics and Security **9**(7), 1084–1097 (2014)
7. Fierrez, J., Pozo, A., Martinez-Diaz, M., Galbally, J., Morales, A.: Benchmarking Touchscreen Biometrics for Mobile Authentication. IEEE Trans. on Information Forensics and Security **13**(11), 2720–2733 (2018)
8. Fierrez, J., Morales, A., Vera-Rodriguez, R., Camacho, D.: Multiple classifiers in biometrics. part 2: Trends and challenges. Information Fusion **44**, 103–112 (2018)
9. Galbally, J., Gomez-Barrero, M., Ross, A.: Accuracy evaluation of handwritten signature verification: Rethinking the random-skilled forgeries dichotomy. In: IEEE Int. Joint Conf. on Biometrics (IJCB). pp. 302–310 (2017)
10. Hadid, A., Evans, N., Marcel, S., Fierrez, J.: Biometrics systems under spoofing attack: an evaluation methodology and lessons learned. IEEE Signal Processing Magazine **32**(5), 20–30 (2015)
11. Hernandez-Ortega, J., Fierrez, J., Morales, A., Tome, P.: Time Analysis of Pulse-based Face Anti-Spoofing in Visible and NIR. In: IEEE CVPR Computer Society Workshop on Biometrics (2018)
12. Li, X., Komulainen, J., Zhao, G., Yuen, P.C., Pietikäinen, M.: Generalized face anti-spoofing by detecting pulse from face videos. In: Int. Conf. on Pattern Recognition (ICPR). pp. 4244–4249. IEEE (2016)
13. Mahmoud, T.M., et al.: A new fast skin color detection technique. World Academy of Science, Engineering and Technology **43**, 501–505 (2008)
14. Marcel, S., Nixon, M.S., Fierrez, J., Evans, N.: Handbook of Biometric Anti-Spoofing. 2nd Edition. Springer (2018)
15. McDuff, D., Gontarek, S., Picard, R.W.: Improvements in remote cardiopulmonary measurement using a five band digital camera. IEEE Trans. on Biomedical Engineering **61**(10), 2593–2601 (2014)
16. Perera, P., Patel, V.M.: Efficient and low latency detection of intruders in mobile active authentication. IEEE Trans. on Inf. Forensics and Sec. **13**(6) (2018)
17. Poh, M.Z., McDuff, D.J., Picard, R.W.: Advancements in noncontact, multiparameter physiological measurements using a webcam. IEEE Trans. on Biomedical Engineering **58**(1), 7–11 (2011)
18. Rapczynski, M., Werner, P., Al-Hamadi, A.: Continuous low latency heart rate estimation from painful faces in real time. In: Int. Conf. on Pattern Recognition (ICPR). pp. 1165–1170 (2016)
19. Tasli, H.E., Gudi, A., den Uyl, M.: Remote PPG based vital sign measurement using adaptive facial regions. In: Proc. IEEE Int. Conf. on Image Processing (ICIP). pp. 1410–1414 (2014)