Generalized face anti-spoofing by detecting pulse from face videos

Xiaobai Li[1], Jukka Komulainen[1], Guoying Zhao*[1], Pong-Chi Yuen[2], and Matti Pietikäinen[1]

[1]CMVS, University of Oulu, Oulu, Finland.
Email: {xiaobai.li, jukka.komulainen, guoying.zhao, mkp}@ee.oulu.fi
[2]Department of Computer Science, HKBU, Hong Kong
Email: pcyuen@comp.hkbu.edu.hk

***Abstract*—Face biometric systems are vulnerable to spoofing attacks. Such attacks can be performed in many ways, including presenting a falsified image, video or 3D mask of a valid user. A widely used approach for differentiating genuine faces from fake ones has been to capture their inherent differences in (2D or 3D) texture using local descriptors. One limitation of these methods is that they may fail if an unseen attack type, e.g. a highly realistic 3D mask which resembles real skin texture, is used in spoofing. Here we propose a robust anti-spoofing method by detecting pulse from face videos. Based on the fact that a pulse signal exists in a real living face but not in any mask or print material, the method could be a generalized solution for face liveness detection. The proposed method is evaluated first on a 3D mask spoofing database 3DMAD to demonstrate its effectiveness in detecting 3D mask attacks. More importantly, our cross-database experiment with high quality REAL-F masks shows that the pulse based method is able to detect even the previously unseen mask type whereas texture based methods fail to generalize beyond the development data. Finally, we propose a robust cascade system combining two complementary attack-specific spoof detectors, i.e. utilize pulse detection against print attacks and color texture analysis against video attacks.**

***Index Terms*—Face liveness, pulse, anti-spoofing, cross-database, mask.**

I. Introduction

Face is one of the most popular biometric traits used in authentication systems [1]. Face authentication systems are known to be vulnerable to presentation attacks because presenting a replica of the targeted face is easy compared with falsifying other biometric traits, *e.g.* fingerprint or iris. First, face biometric data can be widely sampled in public or social media [2]. Second, the attacks can be performed in different and relatively cheap ways using *e.g.* photos or videos, or even (3D) masks of the targeted face. The face anti-spoofing problem has received significant attention lately, and many software-based and hardware-based countermeasures have been proposed. Here, we list several papers closely related to the current work. More comprehensive surveys can be found in *e.g.* [1], [3] and [4].

Assuming that there are inherent disparities between genuine faces and fake ones (*e.g.* printed photos), such as shading, reflectance and skin texture (quality), some earlier works [5]–[7] have proposed approaches for performing spoof detection from (single) static images. The key idea has been to capture these attributes by analyzing *e.g.* the spatial frequency power distribution [5], [6] or local texture [7]. Dynamic methods, *i.e.* exploiting facial motion as a clue, have also been explored for face spoof detection. Typically dynamic countermeasures to photo attacks aim at detecting physiological signs of life, such as eye blinks [8] and mouth movements [9]. Since prints and display devices are flat objects whereas live faces are complex 3D structures, low-cost depth sensors, *e.g.* Microsoft Kinect, can be also exploited to simplify the measurement of three-dimensionality of the observed face [10].

The main focus of previous examples in face anti-spoofing research has been on tackling the problem of photo and video attacks, while wearable (3D) mask attacks have received much less attention. The main reason is that it is expensive to collect large scale datasets of masks. However, due to advances in 3D printing, the manufacturing costs of 3D masks are becoming reasonable, which makes mask attacks a significant threat to face biometric systems in addition to print and video attacks.

Recently, Erdogmus *et al.* [11] addressed this issue by releasing the first 3D mask attack dataset 3DMAD in which the attackers wear 3D facial masks of the targeted persons. They showed also that these kinds of attacks are capable of fooling anti-spoofing methods utilizing eye blinking [8] or depth information [10]. Inspired by [7], Erdogmus *et al.* used local binary pattern (LBP) [12] based facial texture representation for 3D mask attack detection. Although the texture based method performs well on the 3DMAD, one potential limitation is worthy of concern. The masks included in the 3DMAD suffer from obvious 3D printing artifacts (see, Figure 1 left), which can be easily captured using powerful texture descriptors, like LBP. However as manufacturing techniques develop, higher quality 3D masks with realistic skin-like texture can be obtained by intruders (see, Figure 1 right). It is likely that texture based methods can be outwitted with these kinds of masks. Furthermore, the performance of texture based approaches is known to degrade dramatically in unknown operating conditions because the facial texture models are highly dependent on the used input camera and fake face type [3], [13], [14]. In real-world applications, face biometric systems are operating in open environments when unknown

*Corresponding author.

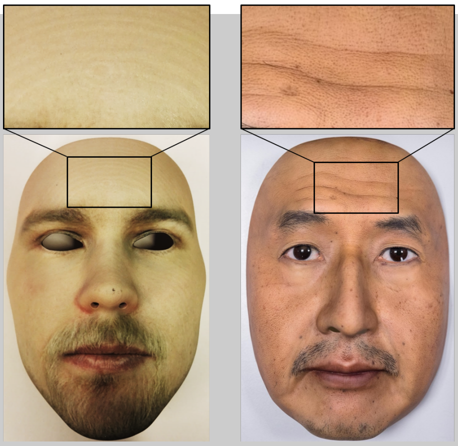

Fig. 1. Comparison of masks used in 3DMAD (left) and REAL-F (right). Upper left: enlarged area highlights the 3D printing defects of the 3DMAD mask. Upper right: enlarged area shows skin-like texture of the REAL-F mask.

input devices and unseen attack scenarios will be definitely encountered. Thus, there is a need for a generalized mask detection approach that does not make too strong assumptions on the input sensor or the mask type, *e.g.* specific texture patterns.

In this present work, we propose to use pulse detection from facial videos for face anti-spoofing. Heart rate measurement from face is an emerging topic that originates from the technique of photoplethysmography (PPG). When light sheds on bare skin parts like earlobes, wrests or fingers, hemoglobins in superficial vessels absorb part of the lights. Cardiac pulse rhythmically changes the number of hemoglobins within a local region, and a PPG can capture the changes by measuring the amount of light being absorbed thus measure the pulse (see, Figure 2). Recent studies have reported that pulse can also be measured from facial videos captured with ordinary color cameras [15], [16]. The background mechanism is similar as the facial skin color changes slightly according to cardiac pulses. These subtle changes can be revealed and used for measuring pulse rate with proper signal processing.

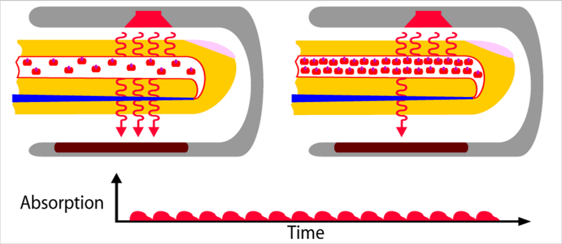

Fig. 2. Illustration[1] of how a photoplethysmography (PPG) works.

Inspired by the pulse measurement studies [15], [16], we analyze these facial color changes that correspond to cardiac pulses in frequency domain, and use their power strengths to build features for the anti-spoofing task. Based on the fact that a pulse signal can be only detected in a real living face but not in any mask (or print) material, the method could be a generalized countermeasure to mask (and print) attacks.

To our best knowledge, this is the first in-depth study that considers pulse detection for the problem of face anti-spoofing. We demonstrate that the pulse detection based method works well on a 3D mask attack database (3DMAD). More importantly, we show that when an unseen mask attack with high quality (*i.e.* REAL-F mask) is included in the test set, texture based methods using LBP features will fail to generalize whereas the pulse detection approach is able to perform robustly. We also explore its effectiveness in detecting print and video replay attacks. Finally, we propose a robust cascade system combining two complementary attack-specific spoof detectors, *i.e.* utilize pulse detection against print attacks and color texture analysis against video attacks. The pulse-based method is very simple yet effective and can be applied in real time. It operates on ordinary color videos, thus no special equipment is needed, which allows it to be generalized to various anti-spoofing scenarios.

For the remaining parts of the paper, we explain the pulse detection method in Section II, and provide the experimental results with discussion in Section III. Our conclusions and plans for future work are presented in Section IV.

II. Method: Pulse Detection from Face Videos for Anti-Spoofing

A. *Face Detection and ROI Tracking*

The method takes a video as the input. Given a facial video of n frames, the first step is to accurately locate and track an area of bare facial skin. We use the lower half face including the cheeks, nose, mouth and chin, while the forehead and eyes are excluded as they may be blocked by glasses or hair.

We apply Viola-Jones face detector [17] on the first frame of the input video, and then use discriminative response map fitting (DRMF) method [18] to find 66 facial landmarks within the face bounding box. We use 9 of the 66 landmarks to define a region of interest (ROI) as shown in Figure 3 a. The location of the ROI is tracked through all frames using the Kanade-Lucas-Tomasi (KLT) algorithm [19].

The tracked ROI contains pixels of facial skin whose color values change with the cardiac pulse. Three raw pulse signals $\boldsymbol{r}_{\text{raw}}$, $\boldsymbol{g}_{\text{raw}}$ and $\boldsymbol{b}_{\text{raw}}$ (see Figure 3 b) are computed one from each RGB channel, respectively. For the red channel, the mean value of all pixels inside the ROI is calculated for each frame, so that the raw pulse signal is a one by n vector $\boldsymbol{r}_{\text{raw}} = [\mathrm{r}_1, \mathrm{r}_2, \ldots, \mathrm{r}_n]$. The two other raw signals $\boldsymbol{g}_{\text{raw}}$ and $\boldsymbol{b}_{\text{raw}}$ are computed similarly.

B. *Temporal Filtering and Power Spectrum Distribution*

Next, we apply three temporal filters that have shown to be helpful in excluding frequencies not relevant for pulse measurement [15]. The first one is a detrending filter based on a smoothness priors approach [20], which is used for reducing slow and non-stationary trend of the signal. The second one is a moving-average filter, which removes random noises by averaging adjacent frames. The third one is a Hamming

[1] Figure from howequipmentworks.com.

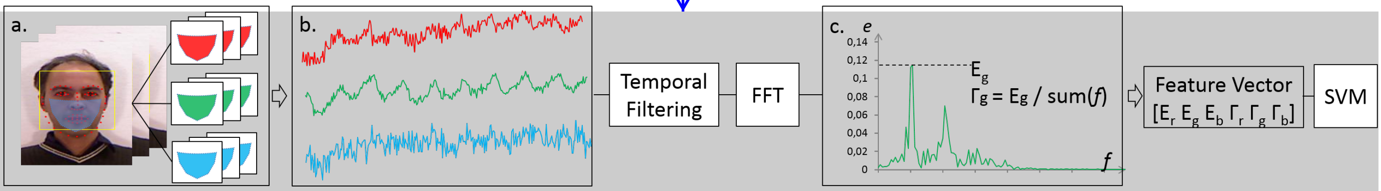

Fig. 3. Framework of the proposed method. For the sake of simplicity, only the PSD of green channel is shown in part c. The PSD curves for red and blue channels are computed in the same way.

window based finite impulse response (FIR) bandpass filter with a cutoff frequency range of $[0.7, 4]$ Hz, which covers the normal range of pulse from 42 beat-per-minute (bpm) to 240 bpm [15].

After pre-processing, we use fast Fourier transform (FFT) to convert the pulse signals into frequency domain. The power spectral density (PSD) curve is computed in which the power e is estimated as a function of the frequency f.

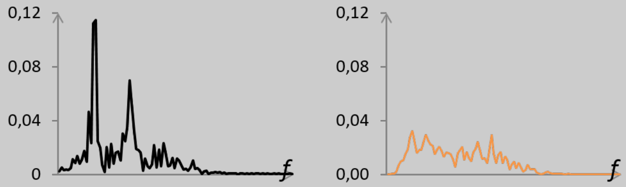

Fig. 4. Typical PSD patterns of a real access (left) and a mask attack (right) extracted from the green color channel.

Figure 4 depicts typical PSD patterns of a genuine access and an attack. If there is a live face in the video, there will be a dominant peak in the PSD corresponding to the pulse frequency and its second harmonic peak. In the case of a mask (or a print) attack, the PSD usually contains just (multiple) random noise peaks at a much lower power level. Therefore, we construct two features for each color channel for the face liveness detection task. The first feature is denoted as E, which is the maximum value of e when f is in the range of $[0.7, 4]$. To increase the stability of the feature for cross-database testing, we build a second feature demoted as Γ, which is the ratio of E and the total power:

$$\Gamma = \frac{E}{\sum_{\forall f \in [0.7,4]} e(f)}. \tag{1}$$

So we have a six-dimensional feature vector $[E_\mathrm{r}, E_\mathrm{g}, E_\mathrm{b}, \Gamma_\mathrm{r}, \Gamma_\mathrm{g}, \Gamma_\mathrm{b}]$ for each video, in which r, g and b indicate corresponding color channels.

C. Classification

The anti-spoofing problem is treated as a two-class classification task, in which real accesses are differentiated from attacks. A support vector machine (SVM) [21] is used as the classifier.

III. Experimental Results and Discussion

We carried out experiments on three datasets in order to evaluate the effectiveness of the pulse-based feature under three different types of attacks: 3D mask, print and video replay attacks. The first two experiments consider two different 3D mask attacks provided in 3DMAD and REAL-F datasets. The 3DMAD is the only publicly available 3D mask attack database, whereas the REAL-F is a new self-collected dataset. In the final experiment, we explore the performance of the porposed approach also in detecting print and video replay attacks provided in the relatively new MSU Mobile Face Spoofing Database (MFSD) [4]. Next, we will describe the experimental set-ups used in our experiments.

Methods*:* For any given dataset, we use the first ten seconds of each video sample (all video samples in the three datasets are at least ten seconds long) to extract a six-dimensional feature vector $[E_\mathrm{r}, E_\mathrm{g}, E_\mathrm{b}, \Gamma_\mathrm{r}, \Gamma_\mathrm{g}, \Gamma_\mathrm{b}]$ (referred to as $Pulse$). In order to mitigate the effect of complex classification schemes and to evaluate the robustness of the proposed feature itself, we used a linear kernel for the SVM with fixed cost parameter $C = 1000$ throughout all experiments.

We considered the widely used LBP features as a baseline for comparing the performance of the pulse-based feature. In the following experiments, the LBP features are first extracted for each frame and the same SVM configuration is used to get image-based classification scores. The video-based performance is obtained by averaging all frame-based score values for each video. Inspired by the state of the art [4], [7], [11], [22], four configurations of LBP features are employed: 1) $LBP-blk$ indicates $\mathrm{LBP}_{8,1}$ histograms extracted from 3×3 blocks of a gray-scale face image and then concatenated into a 531 dimensional vector; 2) $LBP-blk-color$ indicates the same block-wise $\mathrm{LBP}_{8,1}$ but extracted separately from each RGB color channel and then concatenated into a 1593 dimension vector; 3) $LBP-ms$ indicates multi-scale LBP extracted from a whole gray-scale face image combining $\mathrm{LBP}_{8,1}$, $\mathrm{LBP}_{8,2}$, $\mathrm{LBP}_{8,3}$, $\mathrm{LBP}_{8,4}$ and $\mathrm{LBP}_{16,2}$ when the total feature vector length is 479; and 4) $LBP-ms-color$ indicates the same multi-scale LBP but extracted separately from the each RGB channel of a whole face image when the total feature vector length is 1437.

Performance metrics*:* The results for all experiments are reported using equal error rates (EER) which corresponds to the operating point when the false positive rate (FPR) equals the false negative rate (FNR). For the first two experiments, we also report the half total error rate (HTER):

$$HTER = \frac{FPR(\tau^*) + FNR(\tau^*)}{2}, \tag{2}$$

where the threshold τ^* corresponds to the EER operating point of the used development set.

A. 3D Mask Attack Database (3DMAD)

Data: 3DMAD [11] contains 255 videos recorded from 17 subjects. The recording was divided into three sessions s1, s2 and s3: s1 and s2 were real accesses, in which each subject's face was recorded five times (altogether ten videos) at different dates; s3 was the spoof attack condition, in which another person was recorded five times while wearing the 3D face mask of the targeted subject. The masks used in 3DMAD were ordered from online store ThatsMyFace.com and manufactured using 3D printing. A high resolution photo of exactly the same kind of mask is shown in Figure 1 left.

Testing protocol: For training and testing, we use the same leave-one-out cross-validation (LOOCV) protocol as used in paper [11]. In each fold of all 17 folds of cross validations: one subject's data is left for testing, while the data of the remaining 16 subjects is divided into two subject-disjoint halves as training and development sets, respectively.

Results: The results are listed in Table I. It can be seen that the $Pulse$ feature worked well on 3DMAD with an HTER of less than 8% on the test set. Thus, the pulse detection method is very effective in differentiating 3D mask attacks from real access attempts even though a feature vector of only six dimensions and a linear classifier are utilized. On the other hand, all four LBP configurations achieved perfect results on the 3DMAD. However, this can be explained by the obvious 3D printing defects (see Figure 1) that the LBP descriptors can easily capture because both the training set and testing set include data achieved in the same acquisition conditions.

TABLE I
RESULTS ON 3DMAD.

	3DMAD-dev	3DMAD-test	
Method	EER	HTER	EER
$Pulse$	**2.31%**	**7.94%**	**4.71%**
$LBP-blk$	0%	0%	0%
$LBP-blk-color$	0%	0%	0%
$LBP-ms$	0%	0%	0%
$LBP-ms-color$	0%	0%	0%

B. High Quality REAL-F Mask Attack

In this experiment, we demonstrate that the pulse-based approach is able to detect a previously unseen type of mask attack of high quality, while the different texture based methods fail to generalize beyond the development data. As manufacturing techniques improve, higher quality mask are easily accessible. Figure 1 right shows an example of REAL-F mask[2] with an enlarged area highlighting its texture. The highly detailed texture of the REAL-F mask resembles real human skin, thus might be able to spoof texture based countermeasures.

Data: Since there is no other 3D mask database available, we bought two REAL-F masks and collected a small REAL-F dataset[3]. Currently, the REAL-F dataset contains 24 videos of which 12 are real accesses recorded from two subjects, and the other 12 are attacks using two REAL-F masks. All REAL-F videos were recorded using a Logitec C 920 webcam at a frame rate of 30 fps and with a resolution of 1280 by 760. Each video clip lasts ten seconds.

Testing protocol: Again, we randomly select eight subjects' data from 3DMAD for training and the other eight subjects' data as the development set, while the 24 REAL-F videos can be seen as an augmented test set containing a previously unseen mask type. We run 100 folds of such test using different combinations of training data, and summarize the results of all folds to report EER, HTER and the FPRs when FNR=0.1 and FNR=0.01.

Results: Results are shown in Table II. Although the different LBP features performed well on the original 3DMAD data, their performance decreases dramatically in this cross-database scenario, *i.e.* when the REAL-F masks are introduced. From the four types of LBP features, the two block-wise LBP features performed better than multi-scale LBP features. However, even the best-performing configuration ($LBP-blk-color$) misclassifies almost half of the attacks as real accesses when FNR = 0.01. On the other hand, our pulse-based method is able to generalize beyond the development data and performs robustly under the new type of mask attacks.

TABLE II
RESULTS ON REAL-F.

	REAL-F			
Method	HTER	EER	FPR (FNR=0.1)	FPR (FNR=0.01)
$Pulse$	**4.29%**	**1.58%**	**0.25%**	**3.83%**
$LBP-blk$	26.83%	25.08%	37.92%	48.25%
$LBP-blk-color$	25.92%	20.42%	31.50%	48.67%
$LBP-ms$	39.87%	46.50%	59.83%	73.17%
$LBP-ms-color$	47.38%	46.08%	86.50%	95.08%

Face anti-spoofing techniques relying on texture information might be outwitted because of two reasons: 1) they either fail to find texture differences due to the skin-like texture of the high quality mask, 2) or they simply cannot generalize to unseen mask types as they were tuned on a different kind of mask. The pulse detection technique, however, has clear semantic definition and does not make strong assumptions on the mask attack type, *e.g.* specific texture patterns. Therefore, it won't be affected by the mask quality (or new input sensor), and is capable of detecting different kinds of masks, even previously unseen ones, as long as the mask is non-transparent.

Discussion about the pulse-based method for 3D mask attack detection: For further analysis of the pulse feature, the E_{g} values of 255 3DMAD samples and 24 REAL-F samples are depicted in Figure 5 (the other five dimensions have similar patterns). It can be seen that most mask attack cases (orange crosses) are clustered and their E_{g} values are very low. In principle, the PSD of a mask attack corresponds to random noise (see, Figure 4 right), because no pulse power peak can be found in such cases. The E_{g} values for real cases are generally larger because they correspond to pulsation power which is usually presented as a dominant peak with a much higher value than the noise peaks (see, Figure 4 left). The deviation of real

[2] For more information about REAL-F masks, please refer to http://http://real-f.jp/en_news.html.

[3] The dataset will be expanded and made publicly available in near future.

access E_g values, however, is rather high as it can be affected by several factors, including illumination, skin tone, motion and facial resolution [16], [23]. To be more specific, higher facial resolution, *i.e.* amount of skin pixels and reasonable illumination conditions are helpful in order to get stronger pulse signals, whereas darker skin tone or deliberate motion will either derogate the pulse signal or increase noise level.

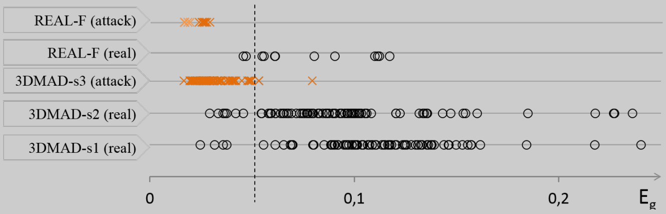

Fig. 5. E_g values of 255 3DMAD samples and 24 REAL-F samples. Black circles show cases of real face. Orange crosses show cases of mask attack.

It is important to notice from Figure 5 that most of the error cases are false negatives, while false positive cases are rare. In general, heart rate cannot be simply detected from masks but in some real cases the pulse signal is hard to reveal due to aforementioned factors. We examined the erroneous cases and found that eight error cases are originated from the same subject who has both darker skin tone and small facial area.

The current REAL-F dataset is somewhat limited but still it shows the limitations of texture based methods and also the potential of our pulse-based method. We believe that these findings will generalize and can be better demonstrated when more comprehensive dataset is available, which will be one important part of our future work.

C. The MSU Mobile Face Spoofing Database (MFSD)

We also explored the effectiveness of the pulse based method in detecting print and video replay attacks. For this purpose, we chose to use the MSU Mobile Face Spoofing Database (MFSD) [4].

Data*:* The MSU MFSD includes print and video attacks. It consists of 280 video clips recorded from 35 subjects using two cameras: built-in camera in MacBook Air 13" (640 by 480), and front-facing camera of the Google Nexus 5 (720 by 480). For each subject, two clips were the real accesses, two clips were photo attacks in which printed HD face photo were held in front of the camera for recording, and four clips were video attacks replayed either by an iPad or by an iPhone.

Testing protocol*:* We divided MSU MFSD data into photo attack (MSU-photo) and video attack (MSU-video) sets to test the proposed method separately on the two different attack types, and then we also test it on the whole dataset (MSU-all). For MSU-photo evaluation, 70 real accesses and 70 print attacks are included; for MSU-video evaluation, 70 real accesses and 140 video attacks are included; for MSU-all evaluation all 280 clips are included. All three tests use the same protocol as proposed in [4], in which 15 subjects' data is used for training and the rest 20 subjects' data for testing. The $Pulse$ feature was extracted for each video the same way as previously done on 3DMAD. For texture features, we considered only $LBP-ms-color$ because it has been suggested in the literature that 1) color LBP features outperform their gray-scale counterparts in print and video attack detection [22], and 2) the use of block-wise LBP is beneficial only in detecting mask attacks [4], [11], [24].

Results*:* The EER results are listed in Table III. It can be seen that the $Pulse$ feature worked best for photo attacks with only 5% EER, but failed for video attacks as expected. For photo attacks, the actual recorded material is paper, thus no pulse power should be detected. On the other hand, for video attacks, the subtle skin color changes caused by pulsation are still present because the presented videos originate from the original live faces (assuming that the quality of the recaptured face is good enough). The pulse-based method cannot differentiate video attacks from real cases as pulse can be detected in both cases. The LBP feature, however, performed better against video attacks than photo attacks.

TABLE III
RESULTS AS EERS ON DIFFERENT SETS OF MSU MFSD.

Method	MSU-photo	MSU-video	MSU-all
$Pulse$	**5.00%**	35.00%	36.67%
$LBP-ms-color$	10.00%	**5.00%**	13.33%
Cascade	–	–	**7.50%**
IDA in [4]	–	–	8.58%*

**Results from [4] are image-based classification results.*

Discussion*:* We did not use the two other well-known face spoofing databases, namely the Replay-Attack [24] and the CASIA [25] datasets, in this experiment because their data is not suitable for the current configuration of the pulse-based feature. As mentioned in the earlier subsection, both facial resolution and motion can affect the pulse signal.

In the Replay-Attack Database, the videos have very low resolution (320 by 240), thus the amount of valid facial skin pixels is too small for recovering pulse signals. According to our prior tests, a face size of 100 by 100 is acceptable for the pulse detection method to work. This is a reasonable requirement considering potential target applications, like mobile biometrics, because nowadays high-quality cameras are embedded in the latest generations of consumer electronics.

The real access videos in the CASIA dataset have fine enough image resolution but they are short and contain a lot of deliberate facial motions, *e.g.* expressions and mouth movements. The subjects were instructed not to keep still because it was argued that facial motion is a crucial liveness cue for anti-spoofing [25]. For the pulse detection method to work properly, the valid part of a input video, *i.e.* with fine facial resolution and without deliberate motion, has to cover at least a couple of circles of heart beats in order to reveal the periodical character of the pulse signal. Due to the short length and the deliberate facial motion, the videos in the CASIA dataset do not meet this criterion.

Some videos in 3DMAD and MSU MFSD also suffer from motion and illumination variations from the pulse detection point of view. In our preliminary experiments, we tested the pulse-based method using videos of different lengths. With longer videos, *e.g.* ten seconds, it achieved the best performance, while with shorter videos, *e.g.*three or five seconds, the method still got reasonable results with only about 1% to 5% increase in the EER and HTER. If all frames of an input video were valid, the pulse method could operate robustly on short video clips, *e.g.* from three to five seconds.

Cascade system*:* Based on previous results, the pulse-based method works better on photo attacks while the LBP method works better on video attacks, but no single feature is robust enough to deal with all types of attacks. We propose a general anti-spoofing system (Figure 6) by cascading the strongest models of the two features. Model 1 was trained on MSU-photo set using the $Pulse$ feature, and Model 2 was trained on MSU-video set using $LBP - ms - color$ feature.

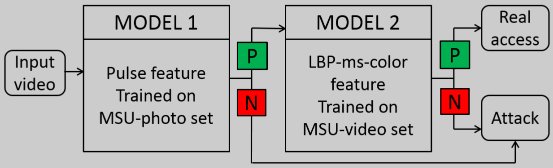
MODEL 1
Input video
Pulse feature Trained on MSU-photo set
P
N
MODEL 2
LBP-ms-color feature Trained on MSU-video set
P
N
Real access
Attack

Fig. 6. Cascaded system combining $Pulse$ and $LBP - ms - color$. ‘P’ indicates positive and ‘N’ negative (classified as real and attack, respectively).

The cascaded system was tested on the whole MSU MFSD. Table III provides a comparison of the proposed system with [4]. For each FNR level of Model 1, we can get an EER for the whole system. The best EER of the cascaded system is 7.5%, achieved when the FNR of Model 1 is set to 2.5%. Since Model 1 and Model 2 are complementary to each other, the cascaded anti-spoofing modules perform robustly under print and video attacks, outperforming the state of the art [4].

IV. Conclusions

In this paper, we proposed a generalized approach for face liveness detection by detecting pulse from facial videos. Power values that correspond to cardiac pulsations are used to build a feature vector for anti-spoofing task. The proposed method was demonstrated to be very effective for detecting 3D mask attacks. While texture-based methods fail to generalize beyond the training data, the proposed method is able to maintain its robust performance even under (previously unseen) high quality mask attacks because the pulse feature is independent from mask quality and type. Our method works well also for print attacks, but is not suitable for video attack detection if used alone. Thus, we proposed a cascaded system combining the pulse feature and color LBP feature that are complementary in detecting both photo and video attacks.

This is the first in-depth study which uses pulse detection for face anti-spoofing. The proposed method has potential to be generalized into many real-world use case scenarios. In future we plan to: 1) expand the REAL-F dataset into a more representative database containing additional high quality 3D mask attacks; 2) explore ways to further increase the robustness of the pulse-based feature in challenging conditions.

Acknowledgment

This work was sponsored by the Academy of Finland, Infotech Oulu and Tekes Fidipro program. This work was also partially supported by Hong Kong RGC General Research Fund HKBU 12201215.

References

[1] J. Galbally, S. Marcel, and J. Fiérrez, “Biometric antispoofing methods: A survey in face recognition,” *IEEE Access*, vol. 2, pp. 1530–1552, 2014.

[2] Y. Li, K. Xu, Q. Yan, Y. Li, and R. H. Deng, “Understanding osn-based facial disclosure against face authentication systems,” in *Proceedings of the 9th ACM Symposium on Information, Computer and Communications Security*, ser. ASIA CCS ’14, 2014, pp. 413–424.

[3] A. Anjos, J. Komulainen, S. Marcel, A. Hadid, and M. Pietikäinen, “Face anti-spoofing: visual approach,” in *Handbook of biometric anti-spoofing*. Springer, 2014, ch. 4, pp. 65–82.

[4] D. Wen, H. Han, and A. Jain, “Face spoof detection with image distortion analysis,” *Information Forensics and Security, IEEE Transactions on*, vol. 10, no. 4, pp. 746–761, April 2015.

[5] J. Li, Y. Wang, T. Tan, and A. K. Jain, “Live face detection based on the analysis of fourier spectra,” pp. 296–303, 2004.

[6] X. Tan, Y. Li, J. Liu, and L. Jiang, “Face liveness detection from a single image with sparse low rank bilinear discriminative model,” in *Computer Vision–ECCV 2010*. Springer, 2010, pp. 504–517.

[7] J. Määttä, A. Hadid, and M. Pietikäinen, “Face spoofing detection from single images using micro-texture analysis,” in *Proceedings of IAPR IEEE International Joint Conference on Biometrics (IJCB)*, 2011.

[8] K. Kollreider, H. Fronthaler, and J. Bigun, “Verifying liveness by multiple experts in face biometrics,” in *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2008, pp. 1200–1205.

[9] K. Kollreider, H. Fronthaler, M. Faraj, and J. Bigun, “Real-time face detection and motion analysis with application in liveness assessment,” *Information Forensics and Security, IEEE Transactions on*, vol. 2, no. 3, pp. 548–558, Sept 2007.

[10] N. Erdogmus and S. Marcel, “Spoofing attacks to 2d face recognition systems with 3d masks,” in *International Conference of the Biometrics Special Interest Group (BIOSIG)*, 2013.

[11] ——, “Spoofing face recognition with 3d masks,” *Information Forensics and Security, IEEE Transactions on*, vol. 9, no. 7, pp. 1084–1097, 2014.

[12] T. Ojala, M. Pietikäinen, and T. Mäenpää, “Multiresolution gray-scale and rotation invariant texture classification with local binary patterns,” *PAMI*, vol. 24, no. 7, pp. 971–987, 2002.

[13] T. de Freitas Pereira, A. Anjos, J. M. De Martino, and S. Marcel, “Can face anti-spoofing countermeasures work in a real world scenario?” in *International Conference on Biometrics*, 2013.

[14] T. Wang, J. Yang, Z. Lei, S. Liao, and S. Z. Li, “Face liveness detection using 3d structure recovered from a single camera,” in *IAPR International Conference on Biometrics, ICB*, June 2013.

[15] M.-Z. Poh, D. J. McDuff, and R. W. Picard, “Advancements in noncontact, multiparameter physiological measurements using a webcam,” *IEEE Trans. on Biomedical Engineering*, 2011.

[16] X. Li, J. Chen, G. Zhao, and M. Pietikainen, “Remote heart rate measurement from face videos under realistic situations,” in *Computer Vision and Pattern Recognition (CVPR)*, 2014, pp. 4264–4271.

[17] P. Viola and M. Jones, “Rapid object detection using a boosted cascade of simple features,” in *CVPR*, 2001, pp. 511–518.

[18] A. Asthana, S. Zafeiriou, S. Cheng, and M. Pantic, “Robust discriminative response map fitting with constrained local models,” in *CVPR*, 2013.

[19] C. Tomasi and T. Kanade, *Detection and Tracking of Point Features*. School of Computer Science, Carnegie Mellon Univ. Pittsburgh, 1991.

[20] M. P. Tarvainen, P. O. Ranta-aho, and P. A. Karjalainen, “An advanced detrending method with application to hrv analysis,” *IEEE Trans. on Biomed. Eng.*, 2002.

[21] C.-C. Chang and C.-J. Lin, “Libsvm: a library for support vector machines,” *TIST*, vol. 2, no. 3, p. 27, 2011.

[22] Z. Boulkenafet, J. Komulainen, and A. Hadid, “Face anti-spoofing based on color texture analysis,” in *Image Processing (ICIP), 2015 IEEE International Conference on*, Sept 2015, pp. 2636–2640.

[23] W. Wang, S. Stuijk, and G. de Haan, “Exploiting spatial redundancy of image sensor for motion robust rppg,” *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 2, pp. 415–425, Feb 2015.

[24] 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 (BIOSIG)*, 2012, pp. 1–7.

[25] Z. Zhang, J. Yan, S. Liu, Z. Lei, D. Yi, and S. Z. Li, “A face antispoofing database with diverse attacks,” in *2012 5th IAPR International Conference on Biometrics (ICB)*, March 2012, pp. 26–31.