Recent Progress on Face Anti-spoofing against 3D Mask Attack

Si-Qi Liu, Pong C. Yuen, Xiaobai Li and Guoying Zhao

```
(filename: chapter13/main.tex)
(corresponding author: Pong C. Yuen)
(authors: Si-Qi Liu, Pong C. Yuen, Xiaobai Li and Guoying
Zhao)
(status: EMPTY [NO])
(status: DRAFT [NO])
(status: COMPLETED [YES])
(action required: to write)
```

Abstract With the advanced 3D reconstruction and printing technologies, creating a super-real 3D facial mask becomes feasible at an affordable cost. This brings a new challenge to face presentation attack detection (PAD) against 3D facial mask attack. As such, there is an urgent need to solve this problem as many face recognition systems have been deployed in real-world applications. Since this is a relatively new research problem, few studies has been conducted and reported. In order to attract more attentions on 3D mask face PAD, this book chapter summarizes the progress in the past few years, as well as publicly available datasets. Finally, some open problems in 3D mask attack are discussed.

Si-Qi Liu and Pong C. Yuen,
Department of Computer Science, Hong Kong Baptist University, Kowloon, Hong Kong, e-mail: {siqiliu,pcyuen}@comp.hkbu.edu.hk

Xiaobai Li and Guoying Zhao
Center for Machine Vision and Signal Analysis, University of Oulu, Finland, e-mail: {xiaobai.li,guoying.zhao}@oulu.fi

1 Background and Motivations

Face presentation attack, a widely used face attack approach where a fake face of an authorized user is present to cheat the face recognition system, is one of the greatest challenges in practice. With the increasing variety of face recognition applications, this security concern has been receiving increasing attentions [13, 15]. Face image or video attacks are the two traditional spoofing methods that can be easily conducted through prints or screens. The face images or videos can also be easily acquired from the internet with the boosting of social networks. In the last decade, a large number of efforts have been devoted to face presentation attack detection (PAD) on photo and video attacks [27, 10, 26, 32, 22, 3, 40, 25, 12, 18, 39, 14, 17, 15, 37] and encouraging results have been obtained.

Recently, with the rapid development of 3D printing and reconstruction technologies, creating a super-real 3D facial mask at an affordable cost become feasible. For instance, to make a customized Thatsmyface 3D mask as shown in Figure 1(a), the user is only required to submit a frontal face image with a few attributed key points. The 3D facial model is reconstructed from it and used to print the 3D mask. Compared with the 2D image or video attacks, 3D masks own 3D structure close to human faces while retaining the appearance in terms of skin texture and facial structure. In this case, the traditional 2D face PAD approaches may not work.

Recent research [9] points out that Thatsmyface[1] masks can successfully spoof many exiting popular face recognition systems. On the 3D Mask Attack Dataset (3DMAD) which is made of Thatsmyface masks, the Inter Session Variability (ISV) modeling method [36] achieves around a Spoofing False Acceptance Rate (SFAR) of around 30%. In addition, the REAL-f mask as shown in Figure 1(b) using the 3D scan and ”Three-Dimension Photo Form (3DPF)” technique to model the 3D structure and print the facial texture, can achieve higher appearance quality and 3D modeling accuracy than the Thatsmyface mask [21]. By observing the detailed textures such as the hair, wrinkle, or even eyes' vessels, without the context information, even a human can hardly identify whether it is a genuine face or not. As such, there is an urgent need to address the 3D mask face PAD.

While 3D mask PAD is a new research problem, some work has been developed and reported which can be mainly categorized as the appearance-based approach, motion-based approach, and remote Photoplethysmography based approach. Similar to the image and video attacks, 3D masks may contain some defects due to the printing quality. Thereby, using the appearance cues such as the texture and color becomes a possible solution [9]. The motion-based approaches detect mask attacks by using the fact that current 3D masks mainly have a hard surface and they cannot retain the subtle facial movements. Motion cues can hardly perform well against 3D mask attacks since 3D masks preserve both the geometric and appearance properties of genuine faces. Moreover, the soft silicone gel mask is able to preserve the subtle movement of the facial skin, which make the motion based approaches less reliable.

[1] `www.thatsmyface.com`

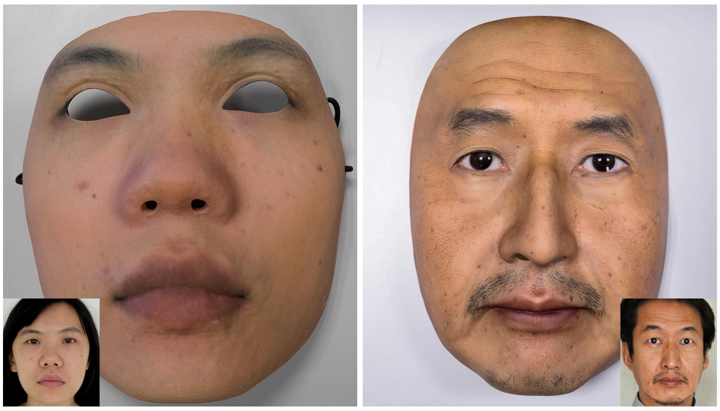

(a) ThatsMyface Mask (b) Real-F Mask

Fig. 1 High resolution sample images of Thatsmyface mask and REAL-f mask.

As such, it is necessary to develop a new intrinsic liveness cue that can be independent of the appearance variation and motion patterns of different masks. Recently, studies turned out that the heartbeat signal on a face can be observed through a normal RBG camera by analyzing the color variation of the facial skin. If it is applied to the 3D mask PAD, the periodic heartbeat signal can be detected on a genuine face but not on a masked face since the mask blocks the light transmission [21]. Due to the uniqueness of this new liveness cues, we categorize is as the remote Photoplethysmography based approach. The book chapter is organized as follows. We have given the background and motivations of this research work. Next, the publicly available 3D mask datasets and evaluation protocols are reviewed in section 2. In section 3, we discuss the methods developed in three categories for face PAD, namely appearance based, motion-based, and remote photoplethysmography based. The performances of the three approaches are evaluated on publicly available datasets in section 4. Finally, open challenges in 3D mask attack are discussed in section 5.

2 Publicly Available Datasets and Experiments Evaluation Protocol

2.1 Datasets

As far as we know, there are two rigid and two soft 3D mask attack datasets for 3D mask PAD. The two rigid mask datasets are: **3D** **M**ask **A**ttack **D**ataset (3D-MAD) [8], **H**ong **K**ong **B**aptist **U**niversity 3D **M**ask **A**ttack with **R**eal World Variation**s** (HKBU-MARs) dataset [20]. The two soft mask datasets are: **S**ilicone **M**ask **A**ttack **D**ataset (SMAD) [23] and **M**ulti-spectral **L**atex Mask based Video **F**ace **P**resentation Attack (MLFP) dataset [2].

2.1.1 3DMAD

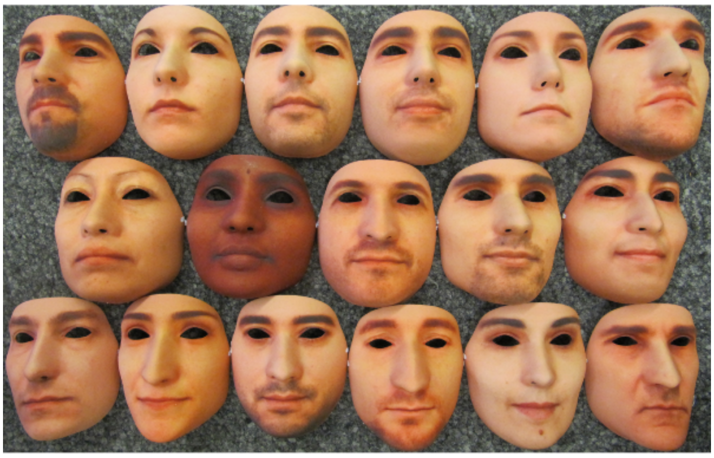


Fig. 2 The 17 customized Thatsmyface masks used in 3DMAD dataset [9].

The **3D** **M**ask **A**ttack **D**ataset (3DMAD) [8] is the first publicly available 3D mask attack dataset in which the attackers wear the customized 3D facial masks of a valid user. The Custom Wearable Masks used in 3DMAD, are built from Thatsmyface.com and proved to be good enough to spoof facial recognition system [8]. The dataset contains a total of 255 videos of 17 subjects, as shown in Figure 2. It is noted that the eyes and nose holes are uncovered by the masks for a better wearing experience. The 3DMAD dataset is recorded by Kinect and contain color and depth information of size 640×480 at 30 frames per second. Each subject has 15 videos with ten live faces and five masked faces. This dataset is divided into three sessions that include two real access sessions recorded with a time delay and one attack session captured by a single operator (attacker). 3D-MAD is the first public available 3D mask dataset which can be downloaded from `https://www.idiap.ch/dataset/3dmad`.

2.1.2 HKBU-MARs

The **H**ong **K**ong **B**aptist **U**niversity 3D **M**ask **A**ttack with **R**eal World Variation**s** (HKBU-MARs) dataset [20] proposes to simulate the real world application scenarios by adding variations in terms of type of mask, camera setting, and lighting condition. In particular, the HKBU-MARs dataset contains 12 subjects with 12 masks. To imitate different types of mask attacks in practice, 6 masks are from Thatsmyface and the other 6 are from REAL-f. Figure 3 shows the masks used in HKBU-MARs. Since face recognition systems are widely deployed in different applications, such as the mobile application and the immigration, 7 different cameras from the stationary and mobile devices are used to capture around 10 seconds videos. For the stationary applications, A web camera Logitech C920, an industrial camera and, and a mirrorless camera (Canon EOS M3) are used to represent different types of face acquisition systems. For the mobile devices, three smartphones: Nexus 5, iPhone 6, Samsung S7 and one tablet: Sony Tablet S are used. In addition, the HKBU-MARs dataset considers 6 lighting conditions to cover the typical scenes of face recognition applications, which includes the Low light, room light, bright light, warm light, side light and upside light as shown in Figure 4. In sum, each subject contains 42 (7 cameras * 6 lightings) genuine and 42 mask sequences and the total size is 1008 videos. The HKBU-MARs that contains the variations of mask type, camera setting, and lighting condition can be used to evaluate the generalization ability of face PAD systems across different application scenarios in practice. More detailed information can be found at `http://rds.comp.hkbu.edu.hk/mars/`. A preliminary version of the HKBU-MARs, namely the HKBU-MARs V1 [21] has been publicly available at `http://rds.comp.hkbu.edu.hk/mars/`. HKBU-MARsV1 contains 8 masks including 6 Thatsmyface masks and 2 REAL-f masks. It is recorded through a Logitech C920 webcam at a resolution of 1280 × 720 under natural lighting conditions.

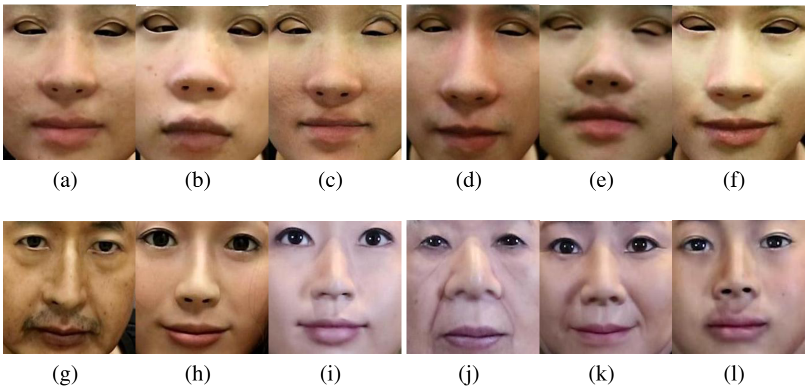


Fig. 3 Sample mask images from the HKBU-MARs dataset. (a)-(f) are ThatsMyFace masks and (g)-(l) are Real-F masks.

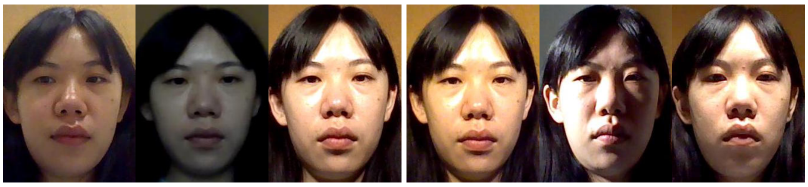


Fig. 4 Different lighting conditions in the HKBU-MARs dataset. (a)-(f) represents the low light, room light, bright light, warm light, side light, and upside light, respectively.

2.1.3 SMAD

Different from 3D printed masks, silicon masks with soft surface can retain both the appearance and facial motions. As shown in the Hollywood movie *Mission Impossible*, Ethan Hunt wears silicon masks to impersonate others identities and even human can hardly recognize them. In this case, face recognition systems are highly vulnerable as well. The **S**ilicone **M**ask **A**ttack **D**atabase (SMAD) [23] is proposed to help research in developing 3D mask attack detection methods in such scenarios. The SMAD contains 130 real and attack videos (65 for each) that obtained from online resources under unconstrained settings. The genuine videos are auditioning, interviewing, or hosting shows collected from multiple sources so they contain different application environment in terms of illumination, background and camera settings. The time interval of videos varies from 3 to 15 second. In particular, the silicon masks in attack videos fit the eyes and mouths holes properly and some masks include hair or mustache. Note that the silicon masks in SMAD are not the customized masks with identities in genuine videos due to the very expensive price. The Text file that contains the videos URLs is available at: `http://www.iab-rubric.org/resources.html`.

2.1.4 MLFP

Since the soft 3D mask can be super real while maintaining the facial movement, using appearance or motion-based methods in the visible spectrum may not be effective. The **M**ulti-spectral **L**atex Mask based Video **F**ace **P**resentation Attack (MLFP) dataset is proposed to help research in designing multi-spectral based face PAD method [2]. The MLFP contains 1350 videos of 10 subjects with or without 7 latex and 3 paper masks. Note that the masks in MLFP are also not the customized masks with identities in genuine videos. The MLFP is recorded in three different spectrums: visible, near-infrared, and thermal with the environmental variations which include indoor and outdoor with fixed and random backgrounds. The MLFP dataset is not yet publicly available at the writing of the chapter.

2.2 Experimental Evaluation Protocol

The performance of a 3D mask face PAD method is mainly evaluated under intra dataset, cross dataset, and intra/cross variation scenarios to test its discriminability and generalizability. The intra dataset testing is conducted in one dataset by separating the subjects into non-overlapping part as the training, development, and testing sets. The cross dataset, and intra/cross variation testing are designed to simulate the scenarios that the training samples are limited and different from the testing samples.

2.2.1 Intra Dataset Testing

For intra dataset testing, the classifier is trained on the training set and test with the testing set. It is noted that the development set is used to tune the parameters for real application scenarios (e.g., the testing set). Erdogmus *et al.* propose to use cross validation to assign different subjects into the training, development and test sets [8]. This protocol is updated to the leave one out cross validation (LOOCV) in [9], which selects one testing subject at each iteration and divides the rest subjects into training and development sets. For instance, the experiments on 3DMAD are done in 17-folds. In each fold, after selecting 1 subject's data as the testing set, for the remaining 16 clients, the first 8 is chosen for training and the remaining for development. Liu *et al.* updated the LOOCV by randomly assigning the remaining subjects for training and development to avoid the effect caused by the order of subjects in a dataset [21].

2.2.2 Cross Dataset Testing

The cross-dataset protocol uses different datasets for training and testing to simulate the practical scenarios where the training data are limited and may be different from the testing samples. To conduct the cross-dataset testing, one can select one dataset for training and use the other dataset for testing. Due to the limited number of subjects of 3D mask attack datasets, the result may not be representative. We may select part of the subjects from one dataset for training and the final result is summarized by conducting several rounds of cross dataset testing. For instance, for the HKBU-MARsV1 and 3DMAD cross testing, Liu *et al.* randomly selected five from the former dataset as the training set and used all data of 3DMAD for testing [21].

2.2.3 Intra Variation Te sting

For the HKBU-MARs dataset [20] that contains three types of variations: mask type, camera, and lighting condition, the intra variation testing protocol is designed to evaluate the robustness of a 3D mask face PAD method when encountering only one specific variation. Under the selected variation (fixed mask type, camera, and lighting condition), LOOCV is conducted to obtain the final results. Although the intra variation testing may not match the practical scenarios, it is useful to evaluate the robustness of a 3D mask PAD method.

2.2.4 Cross-Variation Testing

The cross-variation testing protocol is designed to evaluate the generalization ability across different types of variation in practice. In particular, the leave one variation out cross validation (LOVO) [20] is proposed to evaluate the robustness of one type of variation, the others are fixed. For one specific type of variation, in each iteration, one subvariation is selected as the training set and the rests are regarded as the testing set. For example, for the LOVO on camera type variations, the data captured by one type of camera is selected as the training set and data captured by the rest types of cameras is selected as the testing set. Note that the other types of variation: mask and lighting condition are fixed. In sum, the LOVO of cameras under different mask types and lighting conditions involves a total of 2×6 (mask types × lightings) sets of results [20].

3 Methods

3.1 Appearance based Approach

As the appearance of a face in printed photos or videos is different from the real face, several texture-based methods have been used for face PAD and achieve encouraging results [22, 12, 17, 5]. The 3D mask also contains the quality defect that results in the appearance difference from a genuine face, due to the imperfection precision problems of 3D printing technology. For example, the skin texture and detailed facial structures in masks as shown in Figure 1(a) have perceivable differences compared to those in real faces.

Erdogmus *et al.* evaluate the LBP based methods on 3DMAD dataset and show their effectiveness [9]. The Multi-Scale LBP (MS-LBP) [22] achieves the best performance under most of the testing protocols. From a normalized face image, MS-LBP extracts $LBP^{u2}_{16,2}$, $LBP^{u2}_{8,2}$ from the entire image and $LBP^{u2}_{8,1}$ from the 3×3 overlapping regions. Therefore, one 243-bin, one 59-bin, and nine 59-bin histograms which contain both the global and local information are generated and then concatenated as the final 833-dimensional feature. It is reported that the MS-LBP achieves 99.4% Area Under Curve (AUC), 5% Equal Error Rate (EER) on the Morpho dataset[2]. Multi-Scale LBP [22] concatenates different LBP settings and achieves promising performance on 3D mask detection [9]. While the results are promising

[2] `http://www.morpho.com`

with the above methods, recent studies indicate that they cannot generalize well in a cross-dataset scenario [11, 37]. It is reported that the MS-LBP is less effective (22.6% EER and 86.8% AUC) on HKBU-MARsV1 due to the super-real mask—REAL-f [21].

Since the differences between 3D masks and real faces are mainly from the textures, analyzing the textures details in the frequency domain can be effective. Agarwal *et al.* propose the RDWT-Haralick [1] which uses redundant discrete wavelet transform (RDWT) and Haralick descriptors [16] to analyze the input image under different scales. Specifically, the input image is divided into 3×3 blocks and then the Haralick descriptors are extracted from the 4 sub-bands of the RDWT results and the original image. For video input, the RDWT-Haralick features are extracted from multiple frames and concatenated as the final feature vector. After feature dimension reduction using principal component analysis (PCA), the final result is obtained through Support Vector Machine (SVM) with linear kernel. It is reported that the RDWT-Haralick feature can achieve 100% accuracy and 0% HTER on 3DMAD dataset.

Despite the texture based methods, the perceivable differences between genuine faces and fraud masks also exist in their 3D geometric appearance. Tang *et al.* analyze the 3D meshed faces (acquired through a 3D scanner) and highlight the dissimilarities of micro-shape of genuine and masked faces by the principal curvature measures [7] based 3D geometric attribute [33]. Specifically, they design a shape-description-oriented 3D facial feature description scheme which represents the 3D face as a histogram of principle curvature measures (HOC). It is reported that the HOC can achieve 92% true acceptance rate (TAR) when FAR is 0.01 on Morpho dataset.

Recently, with the booming of the deep learning, the community adopts deep networks to extract discriminative appearance features for biometric PAD [24, 38]. Compared to the solutions that rely on domain knowledge, Menotti *et al.* propose to learn a suitable CNN architecture through the data [24]. In the meantime, the filter weights of the network are learned via back-propagation. The two approaches interact with each other to form the final adapted network, namely the *spoofnet*. The authors report 100% accuracy and 0% HETER on the publicly available 3DMAD dataset. While the performances are significant, the deep learning based features require well designed large-scale training data. Due to the intrinsic data-driven nature [11], the over-fitting problem of the deep learning based methods in cross-dataset scenario remains open.

The hand-crafted features are difficult to be robust across multiple kinds of application scenarios. On the other hand, the deep network based features require significantly large and representative training dataset. Ishan *et al.* propose to use deep dictionary [34] via greedy learning algorithm (DDGL) for PAD [23]. It is reported that DDGL can achieve impressive performance on both the photo, video, hard mask, and silicon mask attacks. On the other hand, its generalizability under cross-dataset scenarios is less promising since the DDGL is also based on the feature learning framework and the performance, to some extent, still depends on the training data.

3.2 Motion based Approach

Since most existing 3D masks are made of hard materials, the facial motion, such as eye blinks and mouth movements, and facial expression changes may not be observed on a masked face. As such, the motion based methods on 2D face PAD, such as the dynamic textures [12] or Histograms of Oriented Optical Flow (HOOF) [6] are effective in detecting these hard 3D masks.

Talha *et al.* propose the multifeature Videolet by encoding the appearance texture with the motion information [30] of both facial and surrounding regions. The texture feature is based on a configuration of local binary pattern, namely the multi-LBP. The motion feature is encoded by extracting HOOF from different time slots of the input video. The multi-feature Videolet method not only achieves 0% EER on the 3DMAD dataset but is also effective in detecting the image and video presentation attacks.

Shao *et al.* propose to exploit the lower convolutional layer to obtain the dynamic information from fine-grained textures in feature channels [28]. In particular, given a preprocessed input video, fine-grained textures in feature channels of a convolutional layer of every frame are first extracted using a pre-trained VGG [31]. Then the dynamic information of textures in each feature channel (of all frames) are estimated using optical flow [4]. To exploit the most discriminative dynamic information, a channel-discriminability constraint is learned through minimizing intra-class variance and maximizing inter-class variance. The authors report 0.56% and 8.85% EER on 3DMAD and HKBU-MARsV1 dataset, respectively. To evaluate the generalizability, the cross-dataset [21] testing is conducted and yields 11.79% and 19.35% EER for 3DMAD to HKBU-MARsV1, and HKBU-MARsV1 to 3DMAD.

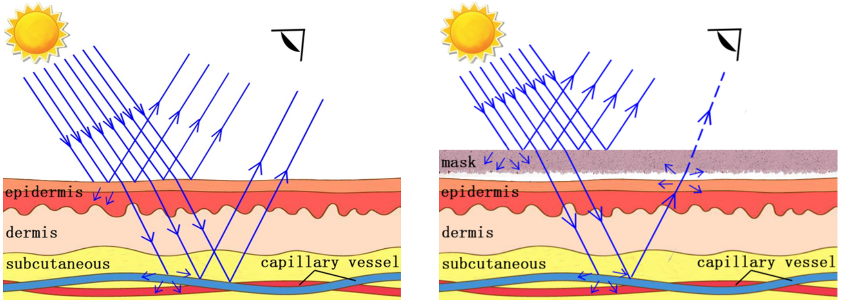
3.3 Remote Photoplethysmography based Approach

Different from the aforementioned two traditional approaches, a new intrinsic liveness cue based on the remote heartbeat detection is proposed recently [21, 19]. The rationale of the rPPG based approach, and two state-of-the-art methods are illustrated in the following subsections.

3.3.1 What is remote Photoplethysmography (rPPG)?

The remote heartbeat detection is conducted through the remote photoplethysmography (rPPG). The photoplethysmography (PPG) is an optically obtained plethysmography, which measures the volumetric of an organ, such as the heart. PPG measures the changes in light absorption of a tissue when heart pumps blood in a cardiac cycle [29]. Different from PPG that often uses the pulse oximeter attached to the skin to detect the signal, rPPG capture it remotely through a normal RGB camera (e.g., web-camera or mobile phone camera) under ambient lighting conditions.

3.3.2 Why rPPG Works for 3D Mask Presentation Attack Detection?



(a) Genuine Face (b) Masked Face

Fig. 5 Effect of rPPG extraction on a genuine face (a), and a masked face (b) [21].

rPPG can be the intrinsic liveness cue for 3D mask face PAD. As shown in Figure 5(a), environmental light penetrates skin and illuminates capillary vessel in the subcutaneous layer. When the heart pumps blood in each cardiac cycle, the blood oxygen saturation changes, which results in a periodic color variation, namely the heartbeat signal. The heartbeat signal then transmits back from the vessels and can be observed by a normal RGB camera. Such an intrinsic liveness cue can be very effective on 3D mask face PAD. As shown in Figure 5(b) for a masked face, the light needs to penetrate the mask and the source heartbeat signal needs to go through the mask again to be observed. Consequently, the rPPG signal extracted from a masked face is too weak to reflect the liveness evidence. In summary, rPPG signals can be detected on genuine faces but not on masked faces, which shows the feasibility of rPPG based 3D mask face PAD.

3.3.3 rPPG-based 3D Mask Presentation Attack Detection

Liu *et al.* are the first that exploit rPPG for 3D mask face PAD [21]. First, the input face is divided into local regions based on the facial landmarks and used to extract local rPPG signals. Then they model a correlation pattern from it to enhance the heartbeat signal and weaken the environmental noise. This is because the local rPPG signals share the same heartbeat frequency and the noise does not. Finally, the local rPPG correlation feature is fed into an SVM tuned by the learned confidence map. The experiments shows that this method achieves 16.2% EER and 91.7% AUC on HKBU-MARsV1 and 95.5% AUC and 9.9% EER on a Combined dataset formed of the 3DMAD and HKBU-MARsV1. In addition, since the rPPG signal is independent of the mask's appearance quality, the local rPPG solution yields good generalizability. Under the cross-dataset testing, it achieves 94.9% and 91.2% AUC for 3DMAD to HKBU-MARsV1 and HKBU-MARsV1 to 3DMAD.

Li *et al.* develop a generalized rPPG based face PAD which works for both 3D mask and traditional image and video attacks [19]. Given the input face video, the Viola-Jones face detector [35] is used to find the bounding box for landmark detection. A customized region of interest is then defined to extract three raw pulse signals from the RGB channels. Next, they apply temporal filters to remove frequencies not relevant for pulse analysis. Signals are analyzed in the frequency domain and the liveness feature, a vector that consists of the maximum power amplitudes and the signal to noise ratio of the three channels, is extracted from the power density curves. This method achieves 4.71% EER on 3DMAD and 1.58% EER on two REAL-f masks [19]. It is also noted that most of the error cases are false negatives. This is because heart rate is fragile due the factors like darker skin tone and small facial resolution.

4 Experiments

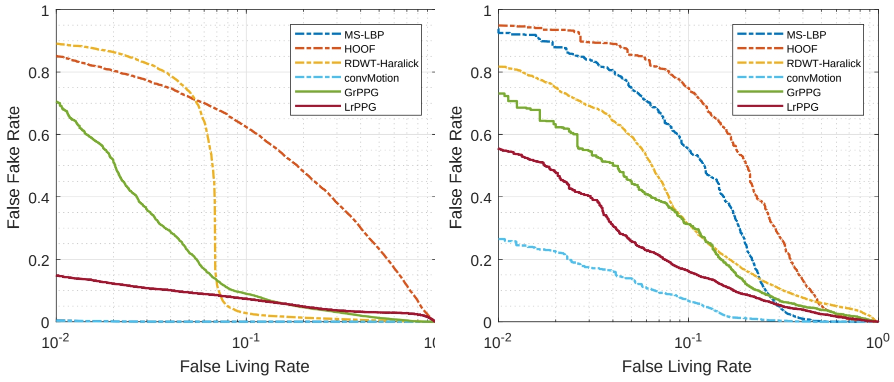
Two experiments are conducted to evaluate the performance of the appearance-based, motion-based, and rPPG-based methods. In particular, MS-LBP [9], HOOF [30], RDWT-Haralick [1], convMotion [28], the local rPPG solution (LrPPG) [21], and the global rPPG solution (GrPPG)[19] are implemented and evaluated under both intra and cross dataset testing protocols. For the appearance based methods, only the first frame of the input video is used. It is also noted that only the HOOF part of the videoLet [30] method is adopted to compare the motion cue with other methods. The final results are obtained through MATLAB SVM with RBF kernel for MS-LBP, HOOF, RDWT-Haralick and LrPPG and linear kernel for GrPPG.

To evaluate the performance, HTER [9, 15], AUC, EER and False Fake Rate (FFR) when False Liveness Rate (FLR) equals 0.1 and 0.01 are used as the evaluation criteria. For the intra-dataset test, HTER is evaluated on the development set and testing set, which are named as HTER_dev and HTER_test, respectively. A ROC curve with FFR and FLR is used for qualitative comparisons.

4.1 Intra Dataset Evaluation

To evaluate the discriminability of the 3D mask methods, LOOCV experiments on both 3DMAD and HKBU-MARsV1 datasets are conducted. For the 3DMAD dataset, we randomly choose 8 subjects for training and the remaining 8 for development. For the HKBU-MARsV1 dataset, we randomly choose 3 subjects as training set and the remaining 4 as the development set. We conduct 20 rounds of experiments.

Table 1 shows the promising results of the texture based method on 3DMAD dataset while the performance drop on the HKBU-MARsV1 points out the limitation on the high-quality 3D masks. HOOF achieves similar results on the two



(a) 3DMAD (b) HKBU-MARsV1

Fig. 6 Average ROC curves of two datasets under intra-dataset protocol.

datasets while the precisions are below the average. Note that the rPPG based methods perform better on 3DMAD than on HKBU-MARsV1. Since the rPPG signal quality depends on the number of pixels of the region of interests, this circumstances may due to the facial resolution of videos from 3DMAD (around 80×80) are smaller than the videos from HKBU-MARsV1 (around 200×200). Specifically, LrPPG achieves better results due to the robustness of the cross-correlation model and confidence map. It is also noted that, as shown in Figure 6 the major error classifications fall on the false reject due to the weakness of rPPG signals on face. The convMotion that fuses deep learning with motion liveness cue achieves the best performance among the existing methods.

Table 1 Comparison of results under intra dataset protocol on the 3DMAD dataset

	HTER_dev(%)	HTER_test(%)	EER(%)	AUC	FFR@ FLR=0.1	FFR@ FLR=0.01
MS-LBP [9]	0.15 ± 0.6	1.56 ± 5.5	0.67	100.0	**0.00**	0.42
HOOF [30]	32.9 ± 6.5	33.8 ± 20.7	34.5	71.9	62.4	85.2
RDWT-Haralick [1]	7.88 ± 5.4	10.0 ± 16.2	7.43	94.0	2.78	89.1
convMotion [28]	**0.10 ± 0.1**	**0.95 ± 0.6**	**0.56**	**100.0**	0.00	**0.30**
GrPPG [19]	8.99 ± 3.1	10.7 ± 11.5	9.41	95.3	8.95	70.7
LrPPG [21]	7.12 ± 4.0	7.60 ± 13.0	8.03	96.2	7.36	14.8

4.2 Cross Dataset Evaluation

To evaluate the generalization ability across different datasets, we design cross-dataset experiments where the training and test samples are from two different

Table 2 Comparison of results under intra dataset protocol on the HKBU-MARsV1 dataset

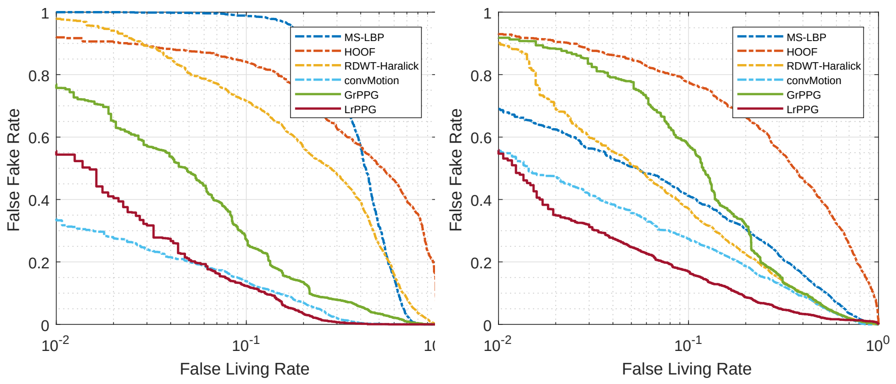
	HTER_dev(%)	HTER_test(%)	EER(%)	AUC	FFR@ FLR=0.1	FFR@ FLR=0.01
MS-LBP [9]	17.5 ± 10.8	11.9 ± 18.2	21.1	86.8	55.4	93.9
HOOF [30]	25.4 ± 12.8	23.8 ± 23.9	28.9	77.9	75.0	95.6
RDWT-Haralick [1]	16.3 ± 6.8	10.3 ± 16.3	18.1	87.3	31.3	81.7
convMotion [28]	**2.31 ± 1.7**	**6.36 ± 3.5**	**8.47**	**98.0**	**6.72**	**27.0**
GrPPG [19]	14.9 ± 6.5	15.8 ± 13.1	16.8	90.2	31.3	73.5
LrPPG [21]	11.5 ± 3.9	13.0 ± 9.9	13.4	93.6	16.2	55.6

datasets. In particular, when 3DMAD and HKBU-MARsV1 are used as the training and testing datasets (3DMAD→HKBU-MARsV1 for short), we randomly select 8 subjects from 3DMAD for training and use the remaining 9 subjects from 3DMAD for development. When HKBU-MARsV1 and 3DMAD are used as the training and testing datasets (HKBU-MARsV1→3DMAD for short), we randomly select 4 subjects from HKBU-MARsV1 as the training set and the remaining 4 subjects from HKBU-MARsV1 as the development set. We also conduct 20 rounds of experiments.

As shown in Table 3 and 4, the rPPG based methods and convMotion achieve close results as the intra dataset testing, which shows their encouraging robustness. The main reason behind the rPPG based method's success is that the rPPG signal for different people under different environment is consistent, so the features of genuine and fake faces can be well separated. On the other hand, the performance of MS-LBP and HOOF drop. Specifically, the appearance based methods, MS-LBP and RDWT-Haralick achieve better results for HKBU-MARsV1→3DMAD, since the HKBU-MARsV1 contains two types of masks while the 3DMAD contains one (the classifier can generalize better when it is trained with larger data variance). It is also noted that the RDWT-Haralick feature achieves better performance than MS-LBP as it analyzes the texture differences from different scales with redundant discrete wavelet transform [1]. The HOOF fails on the cross dataset as for different dataset, the motion patterns based on optical flow may vary with the different recording settings.

Table 3 Cross-dataset evaluation results under 3DMAD→HKBU-MARsV1.

	3DMAD→HKBU-MARsV1				
	HTER(%)	EER(%)	AUC(%)	FFR@ FLR=0.1	FFR@ FLR=0.01
MS-LBP [9]	43.6 ± 5.9	44.8	57.7	98.8	100.0
HOOF [30]	51.8 ± 12.0	50.6	47.3	84.1	92.1
RDWT-Haralick [1]	23.5 ± 4.7	39.6	68.3	71.6	98.1
convMotion [28]	**10.1 ± 2.1**	11.8	96.2	13.8	**33.9**
GrPPG [19]	29.7 ± 11.9	15.6	90.5	27.4	77.0
LrPPG [21]	10.7 ± 3.7	**11.4**	**96.2**	**12.3**	55.6



(a) 3DMAD→HKBU-MARsV1 (b) HKBU-MARsV1→3DMAD

Fig. 7 Average ROC curves under cross-dataset protocol.

Table 4 Cross-dataset evaluation results under HKBU-MARsV1→3DMAD.

	HKBU-MARsV1→3DMAD				
	HTER(%)	EER(%)	AUC(%)	FFR@ FLR=0.1	FFR@ FLR=0.01
MS-LBP [9]	45.5 ± 2.8	25.8	83.4	41.2	69.0
HOOF [30]	42.4 ± 4.1	44.1	57.6	77.6	93.1
RDWT-Haralick [1]	13.8 ± 7.5	21.3	86.7	37.1	90.3
convMotion [28]	17.2 ± 1.3	19.4	89.6	27.4	56.0
GrPPG [19]	29.0 ± 11.6	22.7	83.2	57.3	91.9
LrPPG [21]	**12.8 ± 3.3**	**13.2**	**93.7**	**16.9**	**55.9**

5 Discussion and Open Challenges

With the development of 3D printing and 3D face reconstruction technology, 3D mask is proved to be able to spoof a face recognition system. As such, this problem has drawn an increasing attention with the boosting numbers of related publications. In this chapter, we revealed the challenges of 3D mask presentation attack, summarized the existing datasets and evaluation protocols, and discussed different approaches that have been proposed. Still, there are issues remain open.

Although the costs of 3D masks are expensive, there are two publicly available datasets address the 3D mask presentation attack challenges. However, the numbers of the subjects and customized 3D masks are quite limited because of the cost. Without sufficient number of data, the evaluation results of toy experiments may not be convincing enough for real world applications. Additionally, the variations like the recording device and conditions are limited which results in the difficulties of evaluating the generalization capabilities of the methods in practical scenarios. The existing publicly available datasets using customized masks are mainly recorded through a stationary camera under single lighting condition. While in practice, the training data may vary from the testing samples, in terms of the mask types, camera devices, or lighting conditions. For instance, since the mobile applications are getting more and more popular, the scenario of a mobile device with camera motion interferences under unconstrained lighting conditions could be the common situation. Therefore, more data with real world settings should be designed and collected.

On the other hand, since the 3D mask face PAD is at the beginning stage, current researches mainly focus on the detection under fixed conditions with simple testing protocols, which may not reflect the practical scenarios. The excellent results on single dataset indicate that more challenging evaluation protocols are needed before the 3D mask face PAD can be applied at the practical level. Additionally silicone mask attacks need to be first ”collected” and then studied. For the appearance based methods, adapting different mask qualities, cameras, and lighting conditions are the challenges to be addressed. The motion-based methods may not work on the soft masks that can preserve the facial motion. The rPPG based methods may not be robust under lower lighting condition or with motion interferences. In sum, larger dataset is still the most critical issue for designing more complicated protocols to evaluate not only the discriminability but also the generalizability.

Acknowledgements This project is partially supported by Hong Kong RGC General Research Fund HKBU 12201215, Academy of Finland and FiDiPro program of Tekes (project number: 1849/31/2015).

Index

References

1. Agarwal, A., Singh, R., Vatsa, M.: Face anti-spoofing using haralick features. In: BTAS (2016)
2. Agarwal, A., Yadav, D., Kohli, N., Singh, R., Vatsa, M., Noore, A.: Face presentation attack with latex masks in multispectral videos. SMAD **13**, 130 (2017)
3. Anjos, A., Marcel, S.: Counter-measures to photo attacks in face recognition: a public database and a baseline. In: IJCB (2011)
4. Barron, J.L., Fleet, D.J., Beauchemin, S.S.: Performance of optical flow techniques. International journal of computer vision (1994)
5. Boulkenafet, Z., Komulainen, J., Hadid, A.: Face spoofing detection using colour texture analysis. IEEE Transactions on Information Forensics and Security **11**(8), 1818–1830 (2016)
6. Chaudhry, R., Ravichandran, A., Hager, G., Vidal, R.: Histograms of oriented optical flow and binet-cauchy kernels on nonlinear dynamical systems for the recognition of human actions. In: CVPR (2009)
7. Cohen-Steiner, D., Morvan, J.M.: Restricted delaunay triangulations and normal cycle. In: Proceedings of the nineteenth annual symposium on Computational geometry, pp. 312–321 (2003)
8. Erdogmus, N., Marcel, S.: Spoofing in 2d face recognition with 3d masks and anti-spoofing with kinect. In: BTAS (2013)
9. Erdogmus, N., Marcel, S.: Spoofing face recognition with 3d masks. IEEE Transactions on Information Forensics and Security **9**(7), 1084–1097 (2014)
10. Evans, N.W., Kinnunen, T., Yamagishi, J.: Spoofing and countermeasures for automatic s-peaker verification. In: Interspeech, pp. 925–929 (2013)
11. de Freitas Pereira, T., Anjos, A., De Martino, J.M., Marcel, S.: Can face anti-spoofing countermeasures work in a real world scenario? In: ICB, pp. 1–8. IEEE (2013)
12. de Freitas Pereira, T., Komulainen, J., Anjos, A., De Martino, J.M., Hadid, A., Pietikäinen, M., Marcel, S.: Face liveness detection using dynamic texture. EURASIP Journal on Image and Video Processing **2014**(1), 1–15 (2014)
13. Galbally, J., Marcel, S., Fierrez, J.: Biometric antispoofing methods: A survey in face recognition. IEEE Access **2**, 1530–1552 (2014)
14. Galbally, J., Marcel, S., Fierrez, J.: Image quality assessment for fake biometric detection: Application to iris, fingerprint, and face recognition. IEEE Transactions on Image Processing **23**(2), 710–724 (2014)
15. 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)
16. Haralick, R.M., Shanmugam, K., et al.: Textural features for image classification. IEEE Transactions on systems, man, and cybernetics **1**(6), 610–621 (1973)
17. Kose, N., Dugelay, J.L.: Shape and texture based countermeasure to protect face recognition systems against mask attacks. In: CVPRW (2013)
18. Kose, N., Dugelay, J.L.: Mask spoofing in face recognition and countermeasures. Image and Vision Computing **32**(10), 779–789 (2014)
19. Li, X., Komulainen, J., Zhao, G., Yuen, P.C., Pietikäinen, M.: Generalized face anti-spoofing by detecting pulse from face videos. In: ICPR (2016)
20. Liu, S., Yang, B., Yuen, P.C., Zhao, G.: A 3d mask face anti-spoofing database with real world variations. In: CVPRW (2016)
21. Liu, S., Yuen, P.C., Zhang, S., Zhao, G.: 3d mask face anti-spoofing with remote photoplethysmography. In: ECCV (2016)
22. Määttä, J., Hadid, A., Pietikäinen, M.: Face spoofing detection from single images using micro-texture analysis. In: IJCB (2011)
23. Manjani, I., Tariyal, S., Vatsa, M., Singh, R., Majumdar, A.: Detecting silicone mask based presentation attack via deep dictionary learning. TIFS (2017)
24. Menotti, D., Chiachia, G., Pinto, A., Robson Schwartz, W., Pedrini, H., Xavier Falcao, A., Rocha, A.: Deep representations for iris, face, and fingerprint spoofing detection. Information Forensics and Security, IEEE Transactions on **10**(4), 864–879 (2015)

25. Pan, G., Sun, L., Wu, Z., Lao, S.: Eyeblink-based anti-spoofing in face recognition from a generic webcamera. In: ICCV (2007)
26. Pavlidis, I., Symosek, P.: The imaging issue in an automatic face/disguise detection system. In: Computer Vision Beyond the Visible Spectrum: Methods and Applications (2000)
27. Rattani, A., Poh, N., Ross, A.: Analysis of user-specific score characteristics for spoof biometric attacks. In: CVPRW (2012)
28. Rui Shao, X.L., Yuen, P.C.: Deep convolutional dynamic texture learning with adaptive channel-discriminability for 3d mask face anti-spoofing. In: IJCB (2017)
29. Shelley, K., Shelley, S.: Pulse oximeter waveform: photoelectric plethysmography. Clinical Monitoring, Carol Lake, R. Hines, and C. Blitt, Eds.: WB Saunders Company pp. 420–428 (2001)
30. Siddiqui, T.A., Bharadwaj, S., Dhamecha, T.I., Agarwal, A., Vatsa, M., Singh, R., Ratha, N.: Face anti-spoofing with multifeature videolet aggregation. In: ICPR (2016)
31. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. CoRR **abs/1409.1556** (2014). URL `http://arxiv.org/abs/1409.1556`
32. Tan, X., Li, Y., Liu, J., Jiang, L.: Face liveness detection from a single image with sparse low rank bilinear discriminative model. In: ECCV (2010)
33. Tang, Y., Chen, L.: 3d facial geometric attributes based anti-spoofing approach against mask attacks. In: FG (2017)
34. Tariyal, S., Majumdar, A., Singh, R., Vatsa, M.: Deep dictionary learning. IEEE Access **4**, 10,096–10,109 (2016)
35. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: CVPR (2001)
36. Wallace, R., McLaren, M., McCool, C., Marcel, S.: Inter-session variability modelling and joint factor analysis for face authentication. In: IJCB (2011)
37. Wen, D., Han, H., Jain, A.K.: Face spoof detection with image distortion analysis. IEEE Transactions on Information Forensics and Security **10**(4), 746–761 (2015)
38. Yang, J., Lei, Z., Li, S.Z.: Learn convolutional neural network for face anti-spoofing. arXiv preprint arXiv:1408.5601 (2014)
39. Yi, D., Lei, Z., Zhang, Z., Li, S.Z.: Face anti-spoofing: Multi-spectral approach. In: Handbook of Biometric Anti-Spoofing, pp. 83–102. Springer (2014)
40. Zhang, Z., Yan, J., Liu, S., Lei, Z., Yi, D., Li, S.Z.: A face antispoofing database with diverse attacks. In: ICB (2012)