Pattern Recognition 98 (2020) 107032

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/patcog

A survey on 3D mask presentation attack detection and countermeasures

Shan Jia[a,b], Guodong Guo[c,*], Zhengquan Xu[a,b]

[a] State Key Laboratory of Information Engineering in Surveying Mapping and Remote Sensing, Wuhan University, Wuhan, 430079, China
[b] Collaborative Innovation Center for Geospatial Technology, Wuhan, 430079, China
[c] Lane Department of Computer Science and Electrical Engineering, West Virginia University, Morgantown, WV 26506, USA

ARTICLE INFO

Article history:
Received 23 February 2019
Revised 27 June 2019
Accepted 3 September 2019
Available online 5 September 2019

Keywords:
Face presentation attack
3D Mask spoofing
Biometrics

ABSTRACT

Despite the impressive progress in face recognition, current systems are vulnerable to presentation attacks, which subvert the face recognition systems by presenting a face artifact. Several techniques have been developed to automatically detect different presentation attacks, mostly for 2D photo print and video replay attacks. However, with the development of 3D modeling and printing technologies, 3D mask has become a more effective way to attack the face recognition systems. Over the last decade, various detection methods for 3D mask attacks have been proposed, but there is no survey yet to summarize the advances. We present a comprehensive overview of the state-of-the-art approaches in 3D mask spoofing and anti-spoofing, including existing databases and countermeasures. In addition, we quantitatively compare the performance of different mask spoofing detection methods on a common ground (i.e., using the same database and evaluation metric). The effectiveness of several 2D presentation attack detection methods is also evaluated on two 3D mask spoofing databases to show whether they are applicable or not for 3D mask attacks. Finally, we present some insights and summarize open issues to address in the future.

1. Introduction

Face recognition technologies have been widely used in people's daily lives because of the high efficiency and accuracy. The popularity, however, also makes face recognition systems become a major target of spoofing attacks [1] (also known as presentation attack in ISO/IEC 30107-1). An impostor can fool the biometric system simply by presenting a face artifact of a legitimate user, which can be easily generated due to the easy availability of face images and videos of a person in social networks.

The presented face artifact can be a face of 2D type, using printed/digital photographs or recorded videos on the mobile/tablet, or of 3D type, which is more challenging by wearing a 3D mask on the face [2]. In recent years, the vulnerability of face recognition systems to such spoofing attacks have raised increasing concerns in the academia and industry. Developing presentation attack detection (PAD) methods to determine whether the face at sensor level is real or fake is the efficient countermeasure. Existing methods mainly explore the difference between real faces and face artifacts by a hardware-based analysis (such as sensor characteristics, blink detection, and challenge response) or a software-based detection (including texture, frequency and motion patterns) [3].

Some face PAD methods have achieved promising results in real-world situations, such as phone unlocking applications, mobile payment, security surveillance, and banking services. However, current systems pay more attentions to common 2D face presentation attacks because they are easier and cheaper to implement than 3D presentation attacks. As 3D manufacturing technologies improve, easily attainable facial masks take the presentation attacks one step further and introduce new challenges for PAD studies [4]. The hyper-realistic face masks with 3D structures make it more difficult to tell the difference between the real face and the spoofed one, even for those systems which have already taken spoofing detection into consideration. For example, a young person, disguised himself as an old man using a silicon face and neck mask (see Fig. 1(a)), successfully fooled the border control authorities when boarding a plane from Hong Kong to Canada in 2010 [3]. The Apple's iPhone X, released in 2017, has also been proved by researchers from Bkav that the Face ID can be unlocked when pointed at a face mask of about 200 dollars (see Fig. 1(b)).

Most existing detection methods proposed for fake faces with planar surfaces are rendered futile for 3D masks [4]. For example, texture based methods using the recapture effect of faces in 2D spoofing attacks (with paper-based photos or glass-based video

* Corresponding author.
E-mail address: guodong.guo@mail.wvu.edu (G. Guo).

https://doi.org/10.1016/j.patcog.2019.107032
0031-3203/

(a) (b)

Fig. 1. Examples of 3D mask attacks. (a) Airport security system fooled by silicon mask[1], (b) iPhone X face ID unlocked with 3D mask[2].

screens) may fail to identify 3D mask attacks [5]. The study in [6] shows that a face recognition system with presentation attack detection based on eye blinking and lip movements can also be defeated by photographic masks with eyes and mouth regions cut out. Because the 3D mask generally has more resemblance with human skin, the detection of 3D mask attacks is more challenging and different from traditional 2D PAD methods [7].

Kim [8] first presented a masked fake face detection method using radiance measurements in 2009. After that, a significant amount of literature has been devoted to 3D mask presentation attack detection. However, several surveys [9–11] focused more on the advances in 2D face spoofing detection, but there is no survey yet to summarize and evaluate the recent advances in 3D mask spoofing detection. The aim of this paper is to present a comprehensive overview of the research works in 3D mask spoofing and anti-spoofing. Our main contributions are as follows.

(1) We summarize the characteristics of 3D mask presentation attacks, and review the recent development of 3D mask attack databases and different categories of detection techniques over the past decade. To the best of our knowledge, this paper is the first survey which focuses on 3D mask spoofing and anti-spoofing techniques.
(2) Different 3D mask spoofing detection methods are evaluated and quantitatively compared under a unified framework, i.e., on the same databases, with the same protocols, and using the same metric, to show which kind of methods perform better in detecting 3D mask attacks.
(3) It is usually believed that countermeasures proposed for 2D attacks may fail on the more challenging 3D mask attacks, while we also conduct experiments to evaluate if the 2D PAD methods can still perform well against 3D mask attacks. Surprisingly, experimental results on two public 3D mask attack databases show the outstanding performance and good robustness of some texture based methods.
(4) A comprehensive analysis is provided based on the experimental results, giving some insights into the detection performance, the databases, and some other issues, which helps to have a better understanding of the research in 3D mask spoofing.

The rest of the paper is organized as follows. In Section 2, we introduce the characteristics of 3D mask presentation attacks. Section 3 briefly describes existing 3D mask spoofing databases, and Section 4 reviews the recent research works in 3D mask presentation attack detection. Experimental evaluation of different 3D mask PAD methods and 2D PAD methods against 3D mask attacks are presented in Sections 5. Finally, we provide some insights into the problems and open issues in Section 6, followed by the conclusions in Section 7.

2. 3D Mask presentation attack

In 3D mask presentation attacks, imposers wear masks made of different materials, which have very similar 3D face characteristics to the target face. 3D facial mask spoofing was previously thought impossible to become a common practice in the literature [12], because compared to 2D type attacks, 3D mask is much more difficult and high-cost to manufacture, requiring special 3D devices and materials. However, the recent rapid advancement of 3D printing technologies and services has made it easy and cheap to make hyper-realistic masks. In this section, we briefly introduce existing ways to generate 3D masks.

In the literature, it always requires 3D scanners and printers to generate a mask, no matter by self-manufacturing [13] or relying on third-party services [14,15]. The scanner helps to capture a 3D model of the user face, and then the model is sent to the 3D printer to obtain real size 3D reproduction of the face [13].

Based on different mask materials, the face masks can be classified into hard/rigid and soft/flexible ones [16]. The former can be made of paper, resin, or plastic, which are relatively low-cost. This kind of mask attack is the advanced type of photo-attack in essence; the 3D structure makes it more like a genuine face compared to photo-attack with a 2D planar face. Soft masks, however, often use latex and silicone materials. The flexible surface terrains not only offer closer color and texture fidelity to human skins, but also can adjust to different facial sizes, shapes, and movements, therefore, making the face presentation attack detection more difficult than rigid masks [17].

Importantly, some masks can make the eyes, nostrils, and mouth cavity parts visible via close-fitting holes that match the topology of the face beneath [18]. All these advances in manufacturing 3D masks have made the 3D mask attack more and more popular in practice.

3. 3D Mask attack databases

Due to the advances in 3D mask manufacturing technologies, several 3D mask attack databases have been created to develop new face PAD schemes. In this section, we provide details of 10 existing 3D mask attack databases (see an overview in Table 1 and Fig. 2).

[1] Picture is downloaded from https://chameleonassociates.com/security-breach/.
[2] Picture is downloaded from https://www.theregister.co.uk/2017/11/28/iphone_x_face_id_system_cracked_again/.

Table 1
Summary of existing 3D mask face spoofing databases.

Database	Year	#Subject	#Sample	Material	Sample description
Morpho	2013	20	406	/	2D grayscale images + 3D scans, non-public
3DMAD	2013	17	255	paper, hard resin	2D color images + 2.5D depth maps
3DFS-DB	2016	26	520	plastic	2D, 2.5D images + 3D information, indirect access
BRSU Skin/Face/Spoof	2016	137	141	silicon, plastic, resin, latex	multispectral SWIR, color images
HKBU-MARs	2016	12	1008	/	color images
SMAD	2017	/	130	silicone	color images, from online resources
MLFP	2017	10	1350	latex, paper	visible, NIR, thermal images
ERPA	2017	5	86	resin-coated, silicone	RGB, thermal, NIR images + depth
Rose-Youtu	2018	20	3350	cropped and full paper	color images of 2D and 3D attacks
WMCA	2019	72	1941	plastic, silicone, and paper	multiple channels of 2D and 3D attacks

Morpho Database. Morpho is a non-public mask database, created by MORPHO[3], mainly used for early studies on mask PAD [4,19–22]. This database provides high-quality face samples of 20 subjects, totally 207 real access and 199 mask attack samples, with both the 3D scans, and the corresponding 2D grayscale images.

3D Mask Attack Database (3DMAD). 3DMAD is the first publicly available 3D mask database, proposed in [14]. It comprises 255 video sequences of 17 different users, recorded by an RGB-D camera of Micsoft Kinect device [23] for both real access and presentation attacks using 3D facial masks. The masks are manufactured using services of ThatsMyFace[4] by uploading frontal and profile face images. Two kinds of masks, a life-size wearable mask and a paper-craft mask, are provided for each subject. In all video samples, each frame consists of a depth image, the corresponding color image and manually annotated eye positions.

3D-face spoofing database (3DFS-DB). 3DFS-DB is a self-manufactured and gender-balanced face spoofing database [13]. It consists of two datasets (real and fake) of 26 subjects, 13 men and 13 women. Each dataset contains both videos in.avi format with 2D and 2.5D information, and 3D models in.stl format. The masks are made using two 3D printers: the ShareBot Pro-and the CubeX[5], which are relatively low-cost and worth about 1,000 and 2000 €, respectively. Acrylonitrile Butadiene Styrene (ABS) plastic material is used to generate the physical artifacts. Note that only indirect access to the data is possible for research purposes due to the EU personal data protection regulation, impling that interested

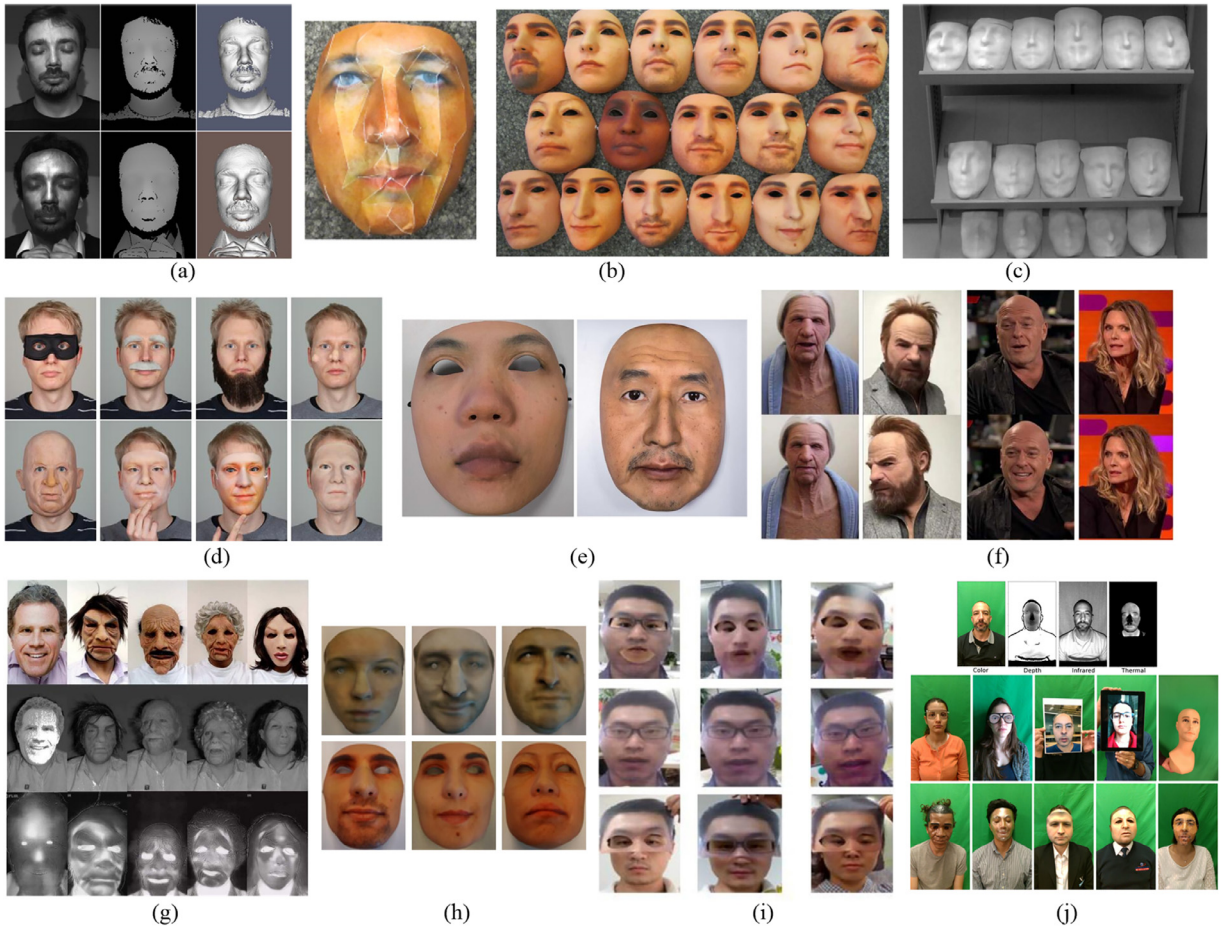


Fig. 2. 3D masks examples in existing databases.[6] (a) Morpho DB, (b) 3DMAD, (c) 3DFS-DB, (d) BRSU Skin/Face/Spoof DB, (e) HKBU-MARs, (f) SMAD, (g) MLFP DB, (h) ERPA DB, (i) Rose-Youtu DB, (j) WMCA.

[3] http://www.morpho.com/.
[4] http://thatsmyface.com/.
[5] https://www.sharebot.it. and http://www.cubify.com.
[6] All pictures are downloaded from the corresponding references.

researchers are allowed to run their algorithms on the database remotely but not to download the data.

BRSU Skin/Face/Spoof Database. This dataset contains 137 subjects and provides multispectral SWIR (Short Wave Infrared) and RGB color images incorporating various types of masks and facial disguises [24]. Two face presentation attack scenarios are considered: disguise of the own identity and counterfeiting of a foreign identity with a mask made of silicon, plastic, latex, or hard resin materials.

HKBU 3D Mask Attack with Real World Variations Database (HKBU-MARs). This is a 3D mask spoofing database with more variations to simulate the real world scenario [15]. It generates 12 masks from two companies (ThatsMyFace and REAL-F[7]) with different appearance qualities. 7 camera types and 6 typical lighting settings are also included to form totally 1008 videos. However, this database only releases version 1 with 110 videos, and the version 2 with 1008 videos is still under construction and not available so far.

Silicone Mask Attack Database (SMAD). This database [25] is collected and compiled from online resources and consists of videos of people wearing silicone masks. It contains 65 genuine access videos of people auditioning, interviewing, or hosting shows, and 65 attacked videos of people wearing a complete 3D structure (but not customized) mask around the head which fits well with proper holes for the eyes and mouth.

Multispectral Latex Mask based Video Face Presentation Attack database (MLFP). MLFP database [16] is a unique multispectral database for face presentation attacks using latex and paper masks. It contains 1350 videos of 10 subjects in visible, near infrared (NIR), and thermal spectrums, which are captured at different locations (indoor and outdoor) in unconstrained environment. Both attack videos and real videos are provided.

ERPA Database. ERPA is a small dataset of bonafide and 3D mask attack presentations, with frame images of 5 subjects stored [17]. The images are captured using two special cameras: the Xenics Gobi thermal camera, and the Intel Realsense SR300 camera recording RGB images, NIR images, and depth information. Both rigid resin-coated masks and flexible silicone masks are considered.

Rose-Youtu Face Liveness Detection Dataset. This dataset [26] is a comprehensive face anti-spoofing database, which covers a large variety of illumination conditions, camera models, and attack types. It consists of 4225 videos of 25 subjects in total, but only 3350 videos of 20 subjects are publically available. Three presentation attack types are created, including printed paper attack, video replay attack, and paper masking attack. For masking attack, the cropped mask, full mask, and upper mask are considered.

Wide Multi Channel Presentation Attack (WMCA) database. This [27] is a multi-channel face presentation attack database with a wide variety of 2D and 3D presentation attacks. It contains 1941 short video recordings of both bonafide and presentation attacks from 72 identities. The data is recorded from different channels, including color, depth, near-infrared and thermal. For 3D mask presentation attacks, it used custom made rigid masks, flexible silicone masks, and paper masks. Additionally, the pulse reading data for bonafide recordings is also provided.

4. 3D Mask PAD methods

Existing 3D mask PAD methods are mainly based on the difference between real face skin and mask materials, which can be classified into the following categories: the reflectance/multispectral properties based, texture based, shape based, deep features based, and other cues/liveness based methods.

[7] http://real-f.jp/en_the-realface.html.

4.1. Reflectance/multispectral properties based detection methods

The earliest studies in 3D mask spoofing detection were based on the reflectance difference of object surfaces. Kim et al. [8] first analyzed the distribution of albedo values for illumination of various wavelengths to find how different facial skins and mask materials (silicon, latex, and skinjell) behave in reflectance, and then created a 2D feature vector that consists of two average radiance values under 850 nm illumination (to distinguish between skins and mask materials) and 685 nm illumination(to distinguish different facial skin colors). Using Fisher's linear discriminant (FLD) classifier, this method achieved 97.78% accuracy in fake face detection on their own experimental data (with mask materials instead of masks). Zhang et al. [7] also measured the albedo curves of skins and non-skin materials, and proposed a distance robust method using two discriminative wavelengths (1450nm and 850nm) to detect fake faces (in photo, video, or mask forms). However, the detection accuracy on mask faces was only 89.18% at multi-distances. Wang et al. [28] pointed out that obtaining images at the band around 1450nm is quite expensive, so they proposed a gradient-based multispectral method using two smaller spectral bands (420nm and 800nm) to detect face presentation attacks. Experiments on their private dataset with planar photos, 3D mannequins, and masks showed promising detection performance, with the True Positive Rate (TPR) of 96.7% and True Negative Rate (TNR) of 97% under the Support Vector Machine (SVM) classifier.

However, all these methods designed PAD schemes independently from the face recognition (FR) process, only focusing on classification performance at the sensor level. Steiner et al. [24] integrated multispectral SWIR skin authentication into existing face verification systems, and also proposed a new public database (named BRSU Skin/Face/Spoof) of corresponding RGB and SWIR images showing different presentation attacks. They designed two methods, one masking out non-skin pixels as a preprocessing step to FR systems, another using a generic region of interest (ROI) as postprocessing of the FR results, and both ensured a high spoof detection performance.

One main limitation of these methods is the requirements of special and expensive devices to acquire multispectral images at various wavelengths. Kose et al. [22,29] tried to obtain the reflectance information from grayscale texture images in the Morpho database without any extra hardware. They used variational retinex algorithm to decompose the texture image into reflectance and illumination components, and then a feature vector was extracted from the reflectance component. Their methods achieved around 95% classification accuracy on their non-public dataset. Table 2 presents a brief overview of these reflectance/multispectral properties based detection methods. Note that only results on 3D mask spoofing databases are reported.

4.2. Texture based detection methods

This kind of methods explore the texture pattern difference of real faces and masks with the help of different texture feature descriptors.

Local Binary Patterns (LBP), one of the most popular face descriptors [36], has been widely used in face presentation attack detection due to its computational simplicity, discriminative power, and robustness to illumination variations. Kose et al. [19] proposed a LBP based method to detect mask attacks, and achieved 88.1% accuracy on Morpho database, slightly better than a depth maps based method (86.0%). With a score level fusion of features from texture images and depth maps, the accuracy was increased to 93.5% in [20]. Erdogmus and Marcel [4,14] compared various types of LBP operators with different classifiers on the proposed 3DMAD

Table 2
Brief overview of reflectance/multispectral properties based detection methods.

Reference	Year	Technique	Spectrum	Database	Performance (classifier)
Kim et al. [8]	2009	Reflectance disparity based on albedo	685+850 nm	Private data	Accuracy = 96.77% (FLD)
Zhang et al. [7]	2011	Multispectral reflectance using Lambertian model	850+1450 nm	Private data	Accuracy = 89.18% (SVM)
Wang et al. [28]	2013	Gradient-based multispectral analysis	420+800 nm	Private data	TPR = 96.70% (SVM), TNR = 97.00% (SVM)
Kose et al. [22]	2013	Reflectance analysis using variational retinex	/	Morpho	Accuracy = 94.47% (SVM)
Kose et al. [29]	2014	Micro-texture analysis on reflectance images	/	Morpho	Accuracy = 95.98% (SVM)
Steiner et al. [24]	2016	Multispectral SWIR skin authentication	1060 nm	BRSU Skin /Face/Spoof	FRR < 5.00% (SVM), FAR < 7.00% (SVM)

Table 3
Brief overview of texture based detection methods.

Reference	Year	Technique	Database	Performance (classifier)
Kose et al. [19]	2013	Multi-scale LBP on texture images	Morpho	Accuracy = 88.1% (SVM)
		Multi-scale LBP on depth maps		Accuracy = 86.0% (SVM)
Kose et al. [20]	2013	Score level fusion of LBP features on texture images and depth maps	Morpho	Accuracy = 93.5%* (SVM)
Erdogmus et al. [14]	2013	Block-based LBP on 2D images	3DMAD	HTER = 0.95% (LDA)
		Block-based LBP on 2.5D images		HTER = 1.27% (LDA)
Erdogmus et al. [4]	2014	Block-based LBP on 2D images	3DMAD,	HTER = 0.12 ± 0.47%(LDA)
		Block-based LBP on 2.5D images	Morpho	HTER = 3.91 ± 6.04%(LDA)
Raghavendra et al. [30]	2014	Local features + global BSIF on 2D and 2.5D images	3DMAD	HTER = 0.03% (SVM)
Raghavendra et al. [31]	2014	Local and global LBP + BSIF on 2D and 2.5D images	3DMAD	ACER = 4.78% (SVM)
Pinto et al. [32]	2015	Time-spectral features and visual codebooks on 2D images	3DMAD	Accuracy = 96.16%, (nonlinear SVM)
Naveen et al. [33]	2016	Local and global LBP + BSIF on 2D and 2.5D images	3DMAD	HTER = 7.65% (Euclidean distance)
Siddiqui et al. [34]	2016	Multi-scale LBP + HOOF on 2D images	3DMAD	EER = 0% (RBF SVM)
Agarwal et al. [35]	2016	Haralick features with block-wise on 2D images	3DMAD	HTER = 0% (SVM)
Liu et al. [15]	2016	Multi-scale LBP	HKBU-MARs	EER ≈ 50% (RBF SVM)

* Using the best result of score level fusion.

database. Results indicated that the classification of block-based LBP features with the Linear Discriminant Analysis (LDA) gives the best results for both color and depth images.

Combining different features in distinguishing between real faces and 3D masks is also one popular and effective way to improve the detection performance. In [30], the global features using Binarized Statistical Image Features (BSIF), and local features (sharp variation and discontinuities) from periocular and nose regions were extracted from both 2D color and depth images on 3DMAD to detect mask attacks. This scheme achieved a satisfactory performance with the Half Total Error Rate (HTER) of 0.03% using the linear SVM classifier. Similarly, both [31] and [33] proposed mask attack detection schemes based on global and local features (LBP and BSIF), and provided superior performance with the Average Classification Error Rate (ACER) of 4.78% and HTER of 7.65%, respectively, on the 3DMAD database. By combining LBP with motion estimation using the Histogram of Oriented Optical Flow (HOOF) features, the study [34] presented a multifeature evidence aggregation approach for both 2D and 3D face presentation attack detection, which achieved an Equal Error Rate (EER) of 0% on the 3DMAD database.

Also aiming at different presentation attacks (including photos, videos, and 3D masks), Pinto et al. [32] took advantage of noise and artifacts of spoofing samples caused by the manufacture and recapture process, and extracted time-spectral features from the video as low-level feature descriptors, and then used the visual codebook concept to find mid-level feature descriptors. Their method performed well in a variety of scenarios and datasets. Recently, Haralick texture features are also explored in [35], showing a good performance in both 2D and 3D mask spoofing databases (with the HTER of 0% on 3DMAD).

A brief overview of these texture based 3D mask spoofing detection methods is shown in Table 3. Although this kind of methods is easy-to-implement and effective on certain databases, their robustness to different mask spoofing attacks needs further investigation. For example, Liu [15] tested different LBP features on their proposed database (HKBU-MARs) with more variations, and showed the LBP based methods can not generalize well when confronting different mask appearance.

4.3. Shape based detection methods

Shape-based 3D mask PAD methods use shape descriptors or image transformation to extract discriminative features from faces and 3D masks.

Kose et al. [21] extracted the 3D face shape information based on warping parameters (WP), and compared its performance with LBP features on 2D images and 2.5D depth maps to analyze the impact of mask spoofing on face recognition systems. They concluded that the syetem based on 3D shape analysis is the most vulnerable to mask attacks (with the highest successful attacks rate of 91.46%). Tang and Chen [37,38] applied 3D shape analysis based on one popular geometric attribute, named principal curvature measures (PCM), and meshedSIFT-based features [42] to describe the meshed facial surface. This method obtained both high verification rates for real faces and satisfactory performance against mask spoofing attacks on the FRCGv2 database (with only genuine 3D face scans) and Morpho database (with 6.72% EER). However, these methods extracted shape information based on spoofing masks stored in 3D triangle meshes, which may limit their applications. Hamdan et al. [39] combined a mask PAD method with the face recognition system. They used the Angular Radial Transformation (ART) to extract shape features from the RGB images as the input to a Maximum Likelihood (ML) classifier. The detection performance on 3DMAD showed the efficiency in discriminating between real faces and masks, with the HTER of 0.91%. They later presented another face recognition method against mask spoofing attacks, which combined the Legendre Moments Invariants (LMI) decomposition of the RGB image with the LDA projection for feature extraction. Using the ML classifier on the 3DMAD datatbase, the Spoof False Acceptance Rate (SFAR) was significantly reduced from 65% to 0.83%.

Table 4
Brief overview of shape based detection methods.

Reference	Year	Technique	Database	Performance (classifier)
Kose et al. [21]	2013	Warping parameters	Morpho	Successful attacks rate = 91.46% (/)
Tang et al. [37,38]	2016	PCM-meshedSIFT facial features	Morpho, FRCGv2	EER = 6.72% (/)
Hamdan et al. [39]	2017	ART on 2D images	3DMAD	HTER = 0.91% (ML)
Hamdan et al. [40]	2018	Combining LMI with LDA projection on 2D images	3DMAD	SFAR = 0.83% (ML)
Wang et al. [41]	2018	Geometry cues reconstruction using 3DMM + mLBP on 2D images	3DMAD	EER = 2.65% (MLK_SVM) EER = 0% (CNN_Softmax)

Table 5
Brief overview of deep features based detection methods.

Reference	Year	Technique	Database	Performance (classifier)
Menotti et al. [43]	2015	Hyperparameter optimization of network architectures (AO) and learning filter weights via back-propagation (FO)	3DMAD	HTER = 0% (AO_SVM), HTER = 24% (FO_Softmax), HTER = 40% (AO+FO)
Feng et al. [44]	2016	Image quality and motion cues fusion using a hierarchical NN	3DMAD	HTER = 0% (Softmax)
Lecena et al. [45]	2017	Transfer learning using the pre-trained VGG-16 model	3DMAD	HTER = 0% (Softmax)
Manjani et al. [25]	2017	Multilevel deep dictionary learning based	3DMAD, SMAD	$HTER_{3DMAD}$ = 0% (SVM), $HTER_{SMAD}$ = 13.1% (SVM)
Shao et al. [46,47]	2018	Facial motion estimation and deep convolutional dynamic texture learning	3DMAD, SUP	$HTER_{3DMAD}$ = 1.76% (SVM), $HTER_{SUP}$ = 13.44% (SVM)
Liu et al. [48]	2018	Several CNN methods on visible and NIR images	Private data	ACER = 3.19% (/)*

* Using the average ACER on two protocols of the best results.

Besides image transformation based methods, reconstructing geometry cues from 2D images through 3D reconstruction models is another way to gain shape features. Wang et al. [41] used the 3D Morphable Model (3DMM) to reconstruct depth cues from RGB images, and then extracted normal features to represent the geometry differences between real faces and masks. They also combined modified LBP (mLBP) features to describe the texture. Experiments on the 3DMAD database compared both hand-craft features and deep learning features, and showed good detection performance with the EER of 2.65% and 0%, respectively.

Table 4 gives a brief summary of the above shape based detection methods. Different from the reflectance-based or some texture-based detection methods, shape-based PAD schemes can be directly applied to the RGB images, with no need of using special sensors to acquire additional information.

4.4. Deep features based detection methods

In contrast to the traditional hand-crafted features, deep feature based methods trend to have a higher detection accuracy and a better generalization ability.

Menotti et al. [43] investigated two deep representation approaches for detecting spoofing in different bimetric modalities. One is based on the hyperparameter optimization of network architectures (AO), and another focues on learning filter weights via back-propagation (referred to as FO). Experiments for the AO and FO approaches along with their combination (AO+FO) on nine databases showed the detection robustness of convolutional networks. For the 3DMAD database, AO method achieved the lowest HTER of 0%, while FO scheme and AO+FO scheme performed poorly in mask spoofing detection with HTER of 24% and 40%, respectively. Lecena et al. [45] also presented a face PAD network (named FAS-Net) to recognize photo, video or mask attacks. It was based on transfer learning using a pre-trained VGG-16 model architecture except for the top layers, and achieved 0% HTER on the 3DMAD database.

Furthermore, some methods tried to combine deep learning based features with hand-craft features, and achieved outstanding results in mask spoofing detection. Feng et al. [44] fused image quality cues (Shearlet) and motion cues (dense optical flow) using a hierarchical neural network to improve the generalization ability for both 2D and 3D spoofing detection. With a bottleneck feature fusion, this method achieved a HTER of 0% on the 3DMAD database. Shao et al. [46] observed that dynamic facial texture information can robustly reflect the face motion patterns, such as eye blinking, lip movements and other spontaneous local facial muscle movements. They then learned the subtle dynamic information from texture features of deep convolutional layers. Both intra-dataset and cross-dataset evaluation on the 3DMAD and their supplementary (SUP) dataset indicated the efficiency and robustness of the proposed method.

Focusing on 3D mask spoofing in varied environments, Manjani et al. [25] introduced a challenging silicon face mask database (SMAD), and also developed a PAD method based on multilevel deep dictionary learning. Experiments were performed on five databases with both 2D and 3D face spoofing attacks, showing promising results in both intra-database and cross-database experiments with the SVM classifier.

Recently in [48], several CNN architectures were investigated on their own dataset to detect 3D masks from visible and NIR images. Results indicated that the multispectral imaging gained better performance than using visible and NIR images separately, and different CNN models showed different detection abilities. We present a summary of existing deep features based 3D mask spoofing detection methods in Table 5.

4.5. Other cues/liveness based detection methods

There are also methods based on other cues of real faces for liveness detection, such as thermal signatures [17], gaze information [53,55,56], and pulse or heartbeat signals [49–52].

Agarwal et al. [16] found that the thermal imaging spectrum is the most effective in detecting mask presentation attacks for different detection methods on their MLFP database (with videos in visible, near infrared, and thermal spectrums). However, thermal imaging devices are always at a high cost and not as easily available as visible imaging cameras.

Using intrinsic liveness signals to distinguish real faces from masked faces is another effective way. In [53–55], gaze information was extracted for detecting both 2D and 3D spoofing attempts on mobile devices. It achieved 0.07% HTER and 0.14% ACER for 3D mask attacks on their private dataset. However, this method

Table 6
Brief overview of 3D mask PAD methods based on other cues.

Reference	Year	Techniques	Databases	Performance (classifier)
Liu et al. [49]	2016	Heartbeat signal analysis based on local rPPG	COMB, SUP	EER_{COMB} = 9.9%, EER_{SUP} = 16.2%(RBFSVM)
Agarwal et al. [16]	2017	Five algorithms in thermal spectrum	MLFP	EER = 10.8%* (SVM)
Li et al. [50]	2017	Pulse analysis based on global PSD signal (green channel)	3DMAD, REAL-F	$HTER_{3DMAD}$ = 7.94%, $HTER_{REAL}$ = 4.29% (SVM)
Liu et al. [51]	2018	Heartbeat analysis based on rPPG correspondence feature	3DMAD, HKBU-MARs	EER_{3DMAD} = 7.44%, EER_{HKBU} = 4.04% (SVM)
Hernandez et al. [52]	2018	Pulse detection based on rPPG	3DMAD, HR	EER_{3DMAD} = 22% (SVM)
Ali et al. [53–55]	2018	Gaze information	Private data	ACER = 14% (k-NN)

* Using the best video-based result of the algorithm combining of Redundant Discrete Wavelet Transform (RDWT) and Haralick features.

Table 7
Summary of advantages and limitations of different methods.

Method	Advantages	Limitations
Reflectance/ multispectral based	Good robustness; good generalizability	Requring special lighting devices
Texture based	Simple to implement; high accuracy	Low robustness; depending on image resolution
Shape based	Simple to implement; high accuracy	High computation cost; sensitive to mask qualities
Deep features based	Very high accuracy; good generalizability	Sensitive to database size
Other cues/livenss based	Difficult to spoof; good generalizability	High computational cost; requiring special devices/ user collaboration; sensitive to lighting and noise

We use 'robustness' to describe the performance stability of detection methods on different databases with the same presentation attacks, while use 'generalizability' to describe the performance of the method on different types of attacks.

is a challenge-response mechanism and requires the user collaboration for capturing eye movements. Photoplethysmography (PPG), as one general way for heart rate monitoring, has been applied for 3D mask spoofing detection in recent years because it can be detected in a non-contacting way using the remote Photoplethysmography technique (rPPG) through a web camera. Liu et al. [49] extracted local rPPG siginals, and showed the detection effectiveness and robustness on two mask spoofing databases (a self-collected Supplementary (SUP) dataset, and a merging database (COMB) of SUP and 3DMAD). To precisely identify the heartbeat information from noisy rPPG signals, they later proposed a liveness feature called rPPG correspondence feature (CFrPPG) in [51], which constructed the correspondence between learned rPPG spectrum templates and local rPPG signals. Although results of this method on the 3DMAD database were slightly worse than some appearance based methods, the performance on HKBU-MARS with hyper real masks (with the EER of 4.04%) was significantly better than other methods (with the EER between 9% and 23%).

Similarly, Li et al. [50] detected pulse signals based on rPPG from facial videos for anti-spoofing. They extracted the global power spectral density (PSD) signal (of green channel) from faces, and quantified it using the maximum value of the spectrum. Experiments on 3DMAD and high quality REAL-F masks datasets demonstrated its effectiveness in detecting 3D mask attacks. Different from Li et al. [50], a smaller and more robust face region (only nose and cheeks) was seleted for rPPG signal extraction in [52]. They also collected a dataset of photo attacks called Heart Rate Database (HR), which contains long videos in visible and NIR spectrum. In time-variant attack scenarios on both 3DMAD and HR databases, experiments showed that longer video sequences resulted in more robust rPPG signals and better detection performance, but the EER was not so satisfactory with more than 20%.

Table 6 presents a brief overview of the liveness cues based detection methods. Generally, this kind of methods performs well in distinguishing real faces from masks. However, extension and application of these methods still have some limitations. For example, thermal signature based methods require special and expensive thermal cameras, while pulse/heartbeat based approaches using remote photo-plethysmography (rPPG) are highly dependent on good light conditions and sensitive to different camera settings (e.g. exposure rates) [46].

For these different categories of methods, we summarize their advantages and limitations in Table 7.

5. Experimental evaluation

In this section, we present a series of experimental comparisons and evaluation on 3D mask attack databases, and try to investigate the following two questions.

(1) Which 3D mask spoofing detection methods perform better based on the same database and evaluation metric?

(2) Will the countermeasures with an outstanding performance for 2D attacks still perform well on 3D mask presentation attacks?

5.1. Comparison results on a common ground

For the first question, we collected several 3D mask attack detection algorithms and evaluated their performance on a common framework, i.e., on the same databases, with the same protocols, and using the same classifier and evaluation metric. Considering the challenge in re-implementing various 3D mask spoofing detection methods, especially some reflectance analysis based and liveness based methods (most relying on special hardwares or with no original codes available for the public), we limit the performance evaluation to some software based approaches with original codes or codes provided by the third party. To the end, 10 state-of-the-art methods were tested, including two alogorithms using reflectance properties [22,29], six texture based methods (multi-scale LBP [19], block-based LBP [14], LBP+BSIF [31], multi-scale LBP+HOOF [34], time-spectral features and codebooks [32], and Haralick features [35]), one shape based method [40], and one learned features based method [45]. These methods were implemented and run under Matlab R2016b on a Windows 10 system with an Intel(R) Core(TM) i7-7500U CPU, 2.70 GHz with a 16 GB RAM, or Python 2.7 under Ubuntu Linux 16.04 LTS with an Intel(R) Core(TM) i7-6850K CPU, 3.60 GHz × 12.

5.1.1. Databases and protocols

The experiments were performed on two publicly available 3D mask attack databases: the 3DMAD (the most widely-used), and the HKBU-MARs-V1 database (with hyper-real 3D masks). Note that the masks in both databases are user-customized, which are closer to reality applications. Two protocols are employed in existing detection methods on these two databases, including the random partition (RP) protocol which divides the whole database into training and testing subsets, and the leave one out cross validation (LOOCV) protocol which is widely used in databases with

Table 8
Random partition protocols of two 3D mask spoofing databases.

Database	/#Subject			#Video				#Frame/video
	Train	Test	Total	Train	Test	Total	Real:Fake	
3DMAD	11	6	17	165	90	255	1:2	300
HKBU-MARs-V1	5	3	8	75	35	110	1:2	300

The training and testing sets have non-overlaps in both subjects and video samples.

Table 9
The HTERs (%) of different 3D methods on mask spoofing databases.

No.	Method	3DMAD		HKBU-MARs-V1		Category
		RP	LOOCV	RP	LOOCV	
A01	Reflectance analysis [22]	**0**	1.18 ± 4.9	37.50	33.57 ± 16.3	Reflectance
A02	Micro-texture and reflectance [29]	**0**	1.18 ± 4.9	31.67	8.57 ± 15.7	Reflectance
A03	Multi-scale LBP [14]	**0**	**0.00 ± 0.0**	**5.83**	5.71 ± 15.1	Texture
A04	Block-based LBP [14]	16.67	4.41 ± 13.9	45.83	17.14 ± 29.3	Texture
A05	LBP+BSIF [31]	**0**	2.94 ± 8.5	58.33	18.57 ± 32.9	Texture
A06	Multi-scale LBP+HOOF [34]	4.17	1.18 ± 4.9	31.67	7.86 ± 15.2	Texture
A07	Time-spectral features and codebooks [32]	11.67	13.59 ± 10.2	25.83	36.43 ± 7.5	Texture
A08	Haralick features [35]	**0**	**0.00 ± 0.0**	51.67	**2.14 ± 5.7**	Texture
A09	LMI based method [40]	**0**	2.06 ± 8.5	65.83	30.00 ± 37.7	Shape
A10	VGG-16 based [45]	9.67	9.59 ± 11.2	33.67	10.0 ± 22.4	Deep

The best results are indicated in bold.

small subject numbers. Therefore, we evaluated all algorithms under these two protocols respectively to show the detection performance. For the random partition protocol, since the two databases do not contain explicit subsets for training and testing, we randomly divided the subjects into two non-overlapping subsets for performance evaluation, so that the training set and testing set are person-disjoint. The details are shown in Table 8.

5.1.2. Detection process

Both the 3DMAD and HKBU-MARs-V1 databases contain videos of 300 frames, while the 3DMAD database also provides depth images. For a higher detection efficiency, we only used the color images as most algorithms do, and randomly selected 20 frames for spoofing detection. The general steps are as follows. First, the faces in the 20 frames were detected, cropped, and normalized into 64 × 64 pixels, similar to the previous approaches [14,15]. The faces in the 3DMAD database were cropped based on the annotated eye positions, while the faces in the HKBU-MARs database were cropped based on the dlib face detector [57]. Next, different features were extracted from each frame. The aggregated feature vectors were fed into a linear SVM classifier to compute the scores, which were then averaged to obtain the final score for a video. Finally, the HTER metric was calculated to report the detection performance.

5.1.3. Evaluation results

Table 9 shows the experimental results of different 3D mask spoofing detection methods on the 3DMAD and HKBU-MARs-V1 databases.

We can first observe the big performance differences on the two 3D mask spoofing databases. All methods achieved lower HTERs (between 0% and 14%) on the 3DMAD database, and experienced performance degradation on the HKBU-MAR-V1 database (with HTERs in the range of 2% to 70%). Such results are reasonable since the masks in the HKBU-MARs-V1 are more realistic and closer to real faces (as shown in Fig. 3), making it harder to distinguish from real accesses. Specifically, two texture based methods using the multi-scale LBP features [14] and Haralick features [35] showed outstanding detection performance on both databases, with the HTER of 0% on the 3DMAD database under two protocols, and the HTER under 6% on the HKBU-MARs-V1 database under

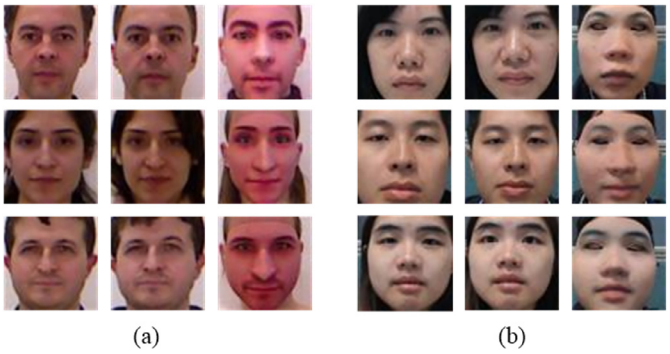


Fig. 3. Examples of cropped faces in two 3D mask spoofing databases. (a) 3DMAD; (b) HKBU-MARs-V1 database. The first two columns are real faces, and the last column is mask spoofing faces.

the LOOCV protocol. This indicates the higher robustness of these two texture feature descriptors. By contrast, the reflectance analysis based method [22], the time-spectral features based method [32], and the LMI decomposition based method [40] demonstrated lower robustness and worse performance in detecting mask attacks in the HKBU-MARs-V1 database, with the HTER over 30%.

Besides, we can see the influence of different protocols on the detection performance. Table 9 shows that most methods achieved lower HTERs under the random partition protocol than the LOOCV protocol on the 3DMAD database, while the results are reversed on the HKBU-MARs-V1 database. This can be attributed to the different subject number of the two databases. The HKBU-MARs-V1 database contains smaller number of subjects (with only 8); therefore, the LOOCV protocol using more training data and longer training time demonstrated a greater advantage in providing more stable detection results than the random partition protocol.

We further explored the influence of classifiers on the 3D mask spoofing detection performance. Four different classifiers (SVM with linear kernel, SVM with RBF kernel, Softmax, and LDA) were used for each 3D mask detection method under the RP protocol on two databases. Optimal parameters were experimentally found with the objective to minimize the error rates on the training set. From the results in Fig. 4(a), we can see that the detection dif-

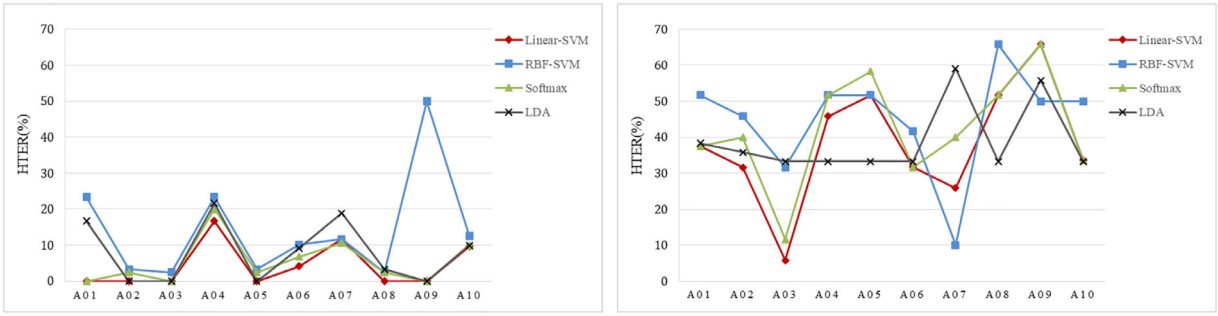


Fig. 4. The HTERs (%) of different methods using different classifiers on two databases. (a) 3DMAD, (b) HKBU-MARs-V1 database.

ferences among the four classifiers are not significant, while the linear SVM classifier achieves the best results for most feature extraction algorithms on the 3DMAD database. For the HKBU-MARs database in Fig. 4(b), however, the HTER values are much higher, and vary greatly using different classifiers. This can be attributed to the highly realistic mask qualities and small data size of this database. Overall, no classifiers work the best for all features, but we can claim that the linear SVM performs better for the LBP based methods (A02–A06), which reaches the similar conclusion as in [4].

5.2. Examining 2D PAD methods for 3D mask attacks

It is usually believed that countermeasures proposed for 2D attacks may fail on 3D mask attacks because of the smaller texture defects, better preserved motion, or more geometric properties in mask attacks [4,29,32,39,49,58]. To verify the correctness, we collected several methods from different categories with outstanding performance against 2D face presentation attacks to evaluate their effectiveness against 3D mask attacks.

5.2.1. Algorithms

Totally there are 10 methods based on different types of features to be evaluated here, namely, (1) motion intensity based [59]; (2) image distortions based [60]; (3) color LBP based [61]; (4) WLD-TOP (Weber Local Descriptor from three orthogonal planes) based [62]; (5) method combining mulit-scale LBP, GLCM (gray level co-ocacurrence matrices) and image distortions [63]; (6) MB-LPQ (multi-block Local Phase Quantization) [64]; (7) color SURF(speeded-up robust features) [65]; (8) method combining LBP and GSLBP (guided scale based LBP) [66]; (9) SqueezeNet features fusing with color LBP [64]; and (10) ResNet-50 based method [67]. For comparison with the results in Section 5.1, we also evaluated these algorithms on the 3DMAD and HKBU-MARs-V1 databases using the HTER metric under the same protocols (the random partition and LOOCV protocol, respectively).

5.2.2. Evaluation results

The experimental results of different 2D PAD methods on the 3DMAD and HKBU-MARs-V1 databases are shown in Table 10.

From Table 10, we can observe the similar performance differences of these methods to the methods in Table 9 on the two mask attack databases. One exception is that two dynamic methods (using motion intensity features [59] and WLD-TOP features [62]) can perform better on the HKBU-MARs-V1 database than on the 3DMAD database. We attribute this to the fact that compared with the 3DMAD database, the cut-out eye regions in the HKBU-MARs-V1 database did not perfectly fit the users' eyes (see Fig. 3). Therefore, the relatively larger eye motion difference between the real accesses and the mask spoofing faces leads to a better detection performance on this database.

Another interesting observation is that several 2D PAD methods, including the texture based and deep learning based, achieved 0% HTERs on the 3DMAD database. Their performance on the more challenging HKBU-MARs-V1 database was also satisfactory compared with the results in Table 9. Specifically, the MB-LPQ based method achieved the best results on the two databases. Other methods using hybrid features, including combining LBP with GSLBP features, fusing multi-scale LBP, GLCM, with image distortions features, and method based on color LBP, also demonstrated an outstanding performance, with the HTERs of 0% on the 3DMAD database, and the average HTER lower than 13% on the HKBU-MARS-V1 database under the LOOCV protocol. This indicates the common perception that spoofing detection methods proposed for 2D attacks will fail on 3D mask attacks is not true. Thanks to the feature robustness, many evaluated face spoofing methods designed for 2D type attacks even showed better performance in detecting 3D mask attacks than methods specially designed for 3D type attacks.

It is also worth noting that two deep learning based methods [64,67] performed worse on the HKBU-MARs-V1 database when compared with their outstanding performance on the 3DMAD database. This is because the deep learning based features are obtained in a data-driven manner, while the HKBU-MARs-V1 database with small data size fail to provide enough data for training the deep models by fine-tuning the pretrained models to their full potential. This also accounts for the similar trend to Table 9 that most methods achieved lower HTER under the LOOCV protocol than the random partition protocol on this database.

5.3. Reported results of existing 3D mask PAD methods

To further show and compare the detection performance of more 3D mask PAD methods, we summarized the reported results from existing detection methods on a unified framework based on the fact that many methods have been evaluated on the same database (3DMAD or HKBU-MARs databases) using the same HTER metric.

The detection performance of 17 methods on the 3DMAD database under two protocols is shown in Table 11. It can be seen that the reported HTERs of most methods are lower than 8% under different classifiers. Deep learning features and texture features show stronger abilities in detecting 3D mask presentation attacks than the liveness cues based methods on this database. Besides, although the 3DMAD database provides both 2D color images and 2.5D depth maps, most algorithms only extract features from 2D color images for computational efficiency and wide generality. Also, the comparison results of block-based LBP method in

Table 10
The HTERs (%) of different 2D methods on 3D mask spoofing databases.

Method	3DMAD		HKBU-MARs-V1		Category
	RP	LOOCV	RP	LOOCV	
Motion intensity [59]	15.00	20.59 ± 14.9	25.83	8.57 ± 22.7	Dynamic
Image distortions [60]	0.83	**0.00 ± 0.0**	42.50	35.71 ± 32.8	Image quality
Color LBP [61]	**0**	**0.00 ± 0.0**	25.00	12.86 ± 18.9	Texture
WLD-TOP [62]	20.83	5.29 ± 12.7	15.00	7.14 ± 12.5	Dynamic
MsLBP+GLCM+distortions [63]	**0**	**0.00 ± 0.0**	31.67	5.71 ± 15.1	Hybrid
MB-LPQ [64]	**0**	**0.00 ± 0.0**	**11.67**	5.71 ± 9.8	Texture
Color SURF [65]	**0**	**0.00 ± 0.0**	60.00	17.14 ± 31.5	Texture
LBP+GSLBP [66]	**0**	**0.00 ± 0.0**	20.00	**2.14 ± 5.7**	Texture
SqueezeNet+color LBP [64]	**0**	**0.00 ± 0.0**	48.33	16.42 ± 28.1	Hybrid
ResNet-50 based [67]	**0**	**0.00 ± 0.0**	20.00	22.14 ± 39.1	Deep

The best results are indicated in bold.

Table 11
Reported results of 3D mask spoofing detection methods on 3DMAD database.

Method	Images	Protocol	HTER(%)	Category
Hyperparameter optimization of network architectures (AO) [43]	2D	RP	**0**	Deep
Transfer learning using the pre-trained VGG-16 model [45]	2D	RP	**0**[*]	Deep
Image quality and motion cues fusion using a hierarchical NN [44]	2D	RP	**0**[*]	Deep
Local features from periocular and nose region+ global BSIF [30]	2D+2.5D	RP	0.03	Texture
Angular Radial Transformation (ART) [39]	2D	RP	0.91[a]	Shape
Block-based LBP [14]	2D	RP	0.95[b]	Texture
Block-based LBP [14]	2.5D	RP	1.27[b]	Texture
Global LBP + BSIF [33]	2D+2.5D	RP	7.65[c]	Texture
Time-spectral features and visual codebooks [32]	2D	RP	8.00	Texture
Multilevel deep dictionary learning based [25]	2D	LOOCV	**0**	Deep
Haralick features with blockwise [35]	2D	LOOCV	**0**	Texture
Facial motion estimation and deep convolutional dynamic texture learning [47]	2D	LOOCV	1.76	Deep
Multi-scale LBP [4]	2D	LOOCV	4.22 ± 10.3[b]	Texture
VGGNet based CNN [68]	2D	LOOCV	6.07 ± 11.3	Deep
Heartbeat analysis based on rPPG correspondence feature [51]	2D	LOOCV	6.82 ± 12.1	Liveness
Pulse analysis based on global PSD signal [50]	2D	LOOCV	7.94	Liveness
Heartbeat signal analysis based on local rPPG [49]	2D	LOOCV	8.57 ± 13.3	Liveness

[*] The best results are indicated in bold. Most results are reported under the SVM classifier, except: [*] using the Softmax classifier;
[a] Using the ML classifier;
[b] Using the LDA classifier;
[c] using the Euclidean distance classifier.

Table 12
Reported results of 3D mask spoofing detection methods on HKBU-MARs database.

Method	Database version	HTER(%)	Category
Heartbeat signal analysis based on local rPPG [49]	V1	**14.70 ± 10.9**[*]	Liveness
Multi-scale LBP [69]	V1	23.00 ± 21.2[*]	Texture
Heartbeat analysis based on rPPG correspondence feature [51]	V2	**4.42 ± 5.1**	Liveness
Heartbeat signal analysis based on local rPPG [49]	V2	8.67 ± 8.8	Liveness
VGGNet based CNN [68]	V2	14.80 ± 22.2	Deep
Pulse analysis based on global PSD signal [50]	V2	16.10 ± 20.5	Liveness
Multi-scale LBP [4]	V2	24.00 ± 25.6	Texture

The best results are indicated in bold. Most results are reported under the linear SVM classifier, except: [*] using the RBF-SVM classifier.

[14] show that extracting features from 2D images yielded better performance, which reached a similar conclusion to [19,33].

Compared with the algorithms in Table 9, we can see four methods [14,32,35,45] in Table 11 were evaluated in Section 5.1. However, three of them [14,32,45] show different performance from our evaluated results. This can be ascribed to the influence of classifiers and ways of database partition in protocol designing.

Table 12 presents the reported results of 6 detection methods under the LOOCV protocol on two versions of HKBU-MARs databases (V1 with 110 videos from 8 subjects while V2 with 1008 videos from 12 subjects). From the limited results shown in Table 12, we can see that liveness based methods using the rPPG signals perform better than the multi-scale LBP and CNN based methods on this superrealistic spoofing database. Further, similarly to the evaluation results in Tables 9 and 10, the HTERs achieved on this database are relatively high, indicating the challenges for existing methods in detecting superrealistic mask attacks.

6. Discussion

Based on the experimental evaluation of different methods in Section 5, we present some insights, and challenges as well, to have a deeper understanding of 3D mask spoofing detection in this section.

6.1. Detection performance

From the detection results in Tables 9–11, first we can see that the detection performance of different methods depends on the database and evaluation protocol. Some deep learning based

and texture features achieved outstanding performance on one database (reaching 0% HTERs on the 3DMAD dataset), but their performance is not satisfactory on another database or under different protocols. Further, some texture based 2D PAD methods have been proved to work well for existing 3D mask databases. This indicates the potential of designing more generalizable countermeasures against various types of presentation attacks. Overall, benefiting from the features' highly discriminative power in local texture description, the methods based on the multi-scale LBP, Haralick features, MB-LPQ, and fusion of LBP with GSLBP features, achieved impressive detection performance against 3D masks attacks in our experimental evaluation.

Inspired by the advances in 2D face PAD methods, the detection performance against 3D mask spoofing attacks could be improved by designing new feature descriptors and exploring new deep learning frameworks [70], or fusing multimodal biometrics [71]. Besides, we emphasize the need to study the interdependency between the 3D mask spoofing detection process and face recognition system. Most existing detection methods for 3D mask spoofing only focus on the PAD performance at the sensor level, without taking the whole recognition process into consideration. Unlike 2D type spoofing, existing 3D mask spoofing databases are more diverse. For some databases with less realistic attacks, the false positive errors of the spoofing detection module may have smaller impact on the false match rate of the face recognition system, while for those databases with hyperreal and user-customized mask attacks, the higher errors (both the false positive and false negative errors) will affect a lot on the face recognition performance. Two recent studies [39,40] presented both the recognition rates and 3D mask spoofing detection performance, which are more complete and powerful to show the effectiveness of the PAD schemes.

6.2. Databases

The databases for 3D mask attacks play a significant role in designing effective and practical detection schemes. Compared to the 2D spoofing samples, 3D mask spoofing is much more complicated and high-cost. Therefore, although several 3D mask databases have been released for the public, there still exist two major issues. One is the lack of large-size 3D mask spoofing databases with more subjects, more types of masks, and more real world variations. This will certainly limit the research works in reporting the anti-spoofing performance. For example, as shown in Table 10, some deep learning based methods showed less advantages on the small HKBU-MARS-V1 than the 3DMAD database. Also, the detection performance of most methods under the common and high-efficient random partition protocol for training and testing is far from satisfactory on small databases.

Another issue is that the different acquisition processes of 3D mask attacks lead to different qualities of spoofing samples, and thus have a great influence on the detection performance. Here we summarize three factors related to the acquisition of 3D mask attack databases, listed as follows.

Mask production process. Existing production of 3D masks is quite diverse, i.e., relying on the third-party services [14,15,17], by self-manufacture based on 3D printers [13], using cut-out papers [26], and using noncustomized masks [16,25]. This process affects a lot on the mask spoofing qualities. For example, the SMAD [25] and MLFP [16] databases used noncustomized masks with textural features such as wrinkles, mustache, beard, and in some cases, facial hair. Compared with the user-customized masks, these masks not only result in larger difference between real accesses and spoofing samples (see Fig. 2), but also make it harder to accurately detect the face regions. Besides, our experimental results showed the influence of the fitness between the cut-out regions and the eyes on the detection performance for motion-based methods.

Sample recording process. Based on the experimental results, we also observed the collection process of mask spoofing videos has an impact on the detection performance. Even using the same 3D mask production service (of ThatsMyFace), the 3DMAD and HKBU-MARs-V1 databases show different spoofing effect with different video recording devices and environment (see Fig. 3). Therefore, the detection results of the same method on these two databases differed greatly. In addition to 2D color images, depth information or multispectral images have been considered using special device in current databases. This helps to provide more discriminative features for 3D mask attack detection, however, the requirement of special devices will restrict its wide application. Besides, multimodal biometric systems have also been explored to protect the systems from presentation attacks, such as fusion of electrocardiogram with face and fingerprint [72], combining face with iris [73], fusing face with fingerprint [71,74]. Although these multimodal schemes require the collection or recording of multiple biometric samples, it is a promising solution to enhance the robustness of biometric systems to different kinds of presentation attacks.

Protocol design. Most existing 3D mask databases do not design protocols for training and testing detection methods. However, from the results in Tables 9 and 10, we can see the large difference of the same detection methods under different protocols, especially in the smaller HKBU-MARs-V1 database. Therefore, the design of proper protocols for different 3D mask databases is necessary to evaluate the effectiveness of the detection schemes.

6.3. Other issues

We also highlight the research of more generalizable algorithms to detect various 3D mask attacks, and also 2D presentation attacks. Most existing 3D mask PAD methods are specifically aimed at 3D mask attacks. It is not clear how they perform for 2D presentation attacks, and it is much harder to learn their detection ability for unknown attacks. Some studies [25,31,34,35,43–45] proposed detection schemes for both 2D and 3D presentation attacks, but few explored the effectiveness of the detection approach in cross dataset experiments. Due to the limited training data and large gap between different types of attacks, the generalizability of existing methods in cross-database experiments still needs to be explored and improved.

Besides, like the detection of 2D presentation attacks, the design of user-friendly 3D mask PAD schemes plays an important role in extending them to real-time applications. Some existing reflectance, texture, or thermal signal based methods require extra hardwares or user collaboration to detect mask spoofing, which are expensive and inconvenient to use in practical applications.

7. Conclusion

The development of 3D mask manufacturing technologies has provided great opportunities for the research on 3D mask spoofing, and PAD techniques as well. In this survey, we have summarized the advances of 3D mask spoofing and anti-spoofing works over the past decade. Over 30 3D mask attack detection methods and 10 3D mask spoofing databases have been analyzed. We have also presented experimental evaluation to quantitatively compare different 3D mask PAD methods under a unified framework, and conducted experiments to show the detection performance of several 2D anti-spoofing methods on 3D mask attack databases. Based on the experimental results, we have presented some insights and open issues, which are beneficial for researchers to develop more effective and generalizable face spoofing detection schemes in the future.

Acknowledgment

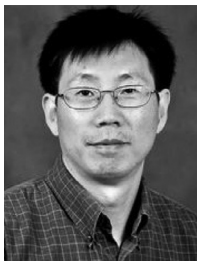
This work was partly supported by an NSF-CITeR project and a WV-HEPC grant, and Applied Basic Research Program of Wuhan (No. 2017010201010114).

References

[1] L. Souza, L. Oliveira, M. Pamplona, J. Papa, How far did we get in face spoofing detection? Eng. Appl. Artif.Intell. 72 (2018) 368–381.
[2] X. Song, X. Zhao, L. Fang, T. Lin, Discriminative representation combinations for accurate face spoofing detection, Pattern Recognit. 85 (2019) 220–231.
[3] R. Ramachandra, C. Busch, Presentation attack detection methods for face recognition systems: a comprehensive survey, ACM Comput. Surv. (CSUR) 50 (1) (2017) 8.
[4] N. Erdogmus, S. Marcel, Spoofing face recognition with 3D masks, IEEE Trans. Inf. ForensicsSecur. 9 (7) (2014) 1084–1097.
[5] S. Tirunagari, N. Poh, D. Windridge, A. Iorliam, N. Suki, A.T. Ho, Detection of face spoofing using visual dynamics, IEEE Trans. Inf. ForensicsSecur. 10 (4) (2015) 762–777.
[6] K. Kollreider, H. Fronthaler, J. Bigun, Verifying liveness by multiple experts in face biometrics, in: Computer Vision and Pattern Recognition Workshops, 2008. CVPRW'08. IEEE Computer Society Conference on, IEEE, 2008, pp. 1–6.
[7] Z. Zhang, D. Yi, Z. Lei, S.Z. Li, Face liveness detection by learning multispectral reflectance distributions, in: Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on, IEEE, 2011, pp. 436–441.
[8] Y. Kim, J. Na, S. Yoon, J. Yi, Masked fake face detection using radiance measurements, JOSA A 26 (4) (2009) 760–766.
[9] J. Galbally, S. Marcel, J. Fierrez, Biometric antispoofing methods: a survey in face recognition, IEEE Access 2 (2014) 1530–1552.
[10] L. Li, P.L. Correia, A. Hadid, Face recognition under spoofing attacks: countermeasures and research directions, IET Biom. 7 (1) (2017) 3–14.
[11] S. Bhattacharjee, A. Mohammadi, A. Anjos, S. Marcel, Recent advances in face presentation attack detection, in: Handbook of Biometric Anti-Spoofing, Springer, 2019, pp. 207–228.
[12] Z. Zhang, J. Yan, S. Liu, Z. Lei, D. Yi, S.Z. Li, A face antispoofing database with diverse attacks, in: Biometrics (ICB), 2012 5th IAPR International Conference on, IEEE, 2012, pp. 26–31.
[13] J. Galbally, R. Satta, Three-dimensional and two-and-a-half-dimensional face recognition spoofing using three-dimensional printed models, IET Biom. 5 (2) (2016) 83–91.
[14] N. Erdogmus, S. Marcel, Spoofing in 2D face recognition with 3D masks and anti-spoofing with Kinect, in: Biometrics: Theory, Applications and Systems (BTAS), 2013 IEEE Sixth International Conference on, IEEE, 2013, pp. 1–6.
[15] S. Liu, B. Yang, P.C. Yuen, G. Zhao, A 3D mask face anti-spoofing database with real world variations, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2016, pp. 100–106.
[16] A. Agarwal, D. Yadav, N. Kohli, R. Singh, M. Vatsa, A. Noore, Face presentation attack with latex masks in multispectral videos, in: Computer Vision and Pattern Recognition Workshops, 2017, pp. 275–283.
[17] S. Bhattacharjee, S. Marcel, What you can't see can help you–extended-range imaging for 3D-mask presentation attack detection, in: Proceedings of the 16th International Conference on Biometrics Special Interest Group., Gesellschaft fuer Informatik eV (GI), 2017. EPFL-CONF-231840
[18] J.G. Sanders, Y. Ueda, K. Minemoto, E. Noyes, S. Yoshikawa, R. Jenkins, Hyper-realistic face masks: a new challenge in person identification, Cognit. Res. 2 (1) (2017) 43.
[19] N. Kose, J.-L. Dugelay, Countermeasure for the protection of face recognition systems against mask attacks, in: Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops on, IEEE, 2013, pp. 1–6.
[20] N. Kose, J.-L. Dugelay, Shape and texture based countermeasure to protect face recognition systems against mask attacks, in: Computer Vision and Pattern Recognition Workshops, 2013 IEEE Conference On, IEEE, 2013, pp. 111–116.
[21] N. Kose, J.-L. Dugelay, On the vulnerability of face recognition systems to spoofing mask attacks, in: Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on, IEEE, 2013, pp. 2357–2361.
[22] N. Kose, J.-L. Dugelay, Reflectance analysis based countermeasure technique to detect face mask attacks, in: Digital Signal Processing (DSP), 2013 18th International Conference on, IEEE, 2013, pp. 1–6.
[23] E. Boutellaa, A. Hadid, M. Bengherabi, S. Ait-Aoudia, On the use of Kinect depth data for identity, gender and ethnicity classification from facial images, Pattern Recognit. Lett. 68 (2015) 270–277.
[24] H. Steiner, A. Kolb, N. Jung, Reliable face anti-spoofing using multispectral SWIR imaging, in: Biometrics (ICB), 2016 International Conference on, IEEE, 2016, pp. 1–8.
[25] I. Manjani, S. Tariyal, M. Vatsa, R. Singh, A. Majumdar, Detecting silicone mask-based presentation attack via deep dictionary learning, IEEE Trans. Inf. ForensicsSecur. 12 (7) (2017) 1713–1723.
[26] H. Li, W. Li, H. Cao, S. Wang, F. Huang, A.C. Kot, Unsupervised domain adaptation for face anti-spoofing, IEEE Trans. Inf. ForensicsSecur. 13 (7) (2018) 1794–1809.
[27] A. George, Z. Mostaani, D. Geissenbuhler, O. Nikisins, A. Anjos, S. Marcel, Biometric face presentation attack detection with multi-channel convolutional neural network, IEEE Trans. Inf. ForensicsSecur. (2019).
[28] Y. Wang, X. Hao, Y. Hou, C. Guo, A new multispectral method for face liveness detection, in: Pattern Recognition (ACPR), 2013 2nd IAPR Asian Conference on, IEEE, 2013, pp. 922–926.
[29] N. Kose, J.-L. Dugelay, Mask spoofing in face recognition and countermeasures, Image Vis. Comput. 32 (10) (2014) 779–789.
[30] R. Raghavendra, C. Busch, Novel presentation attack detection algorithm for face recognition system: application to 3D face mask attack, in: Image Processing (ICIP), 2014 IEEE International Conference on, IEEE, 2014, pp. 323–327.
[31] R. Raghavendra, C. Busch, Robust 2D/3D face mask presentation attack detection scheme by exploring multiple features and comparison score level fusion, in: Information Fusion, 2014 17th International Conference on, IEEE, 2014, pp. 1–7.
[32] A. Pinto, H. Pedrini, W.R. Schwartz, A. Rocha, Face spoofing detection through visual codebooks of spectral temporal cubes, IEEE Trans. Image Process. 24 (12) (2015) 4726–4740.
[33] S. Naveen, R.S. Fathima, R. Moni, Face recognition and authentication using LBP and BSIF mask detection and elimination, in: Communication Systems and Networks, International Conference on, IEEE, 2016, pp. 99–102.
[34] T.A. Siddiqui, S. Bharadwaj, T.I. Dhamecha, A. Agarwal, M. Vatsa, R. Singh, N. Ratha, Face anti-spoofing with multifeature videolet aggregation, in: Pattern Recognition (ICPR), 2016 23rd International Conference on, IEEE, 2016, pp. 1035–1040.
[35] A. Agarwal, R. Singh, M. Vatsa, Face anti-spoofing using Haralick features, in: Biometrics Theory, Applications and Systems (BTAS), 2016 IEEE 8th International Conference on, IEEE, 2016, pp. 1–6.
[36] Z. Lei, M. Pietikäinen, S.Z. Li, Learning discriminant face descriptor, IEEE Trans. Pattern Anal. Mach.Intell. 36 (2) (2014) 289–302.
[37] Y. Tang, L. Chen, Shape analysis based anti-spoofing 3d face recognition with mask attacks, in: International Workshop on Representations, Analysis and Recognition of Shape and Motion FroM Imaging Data, Springer, 2016, pp. 41–55.
[38] Y. Tang, L. Chen, 3D facial geometric attributes based anti-spoofing approach against mask attacks, in: Automatic Face & Gesture Recognition (FG 2017), 2017 12th IEEE International Conference on, IEEE, 2017, pp. 589–595.
[39] B. Hamdan, K. Mokhtar, The detection of spoofing by 3D mask in a 2D identity recognition system, Egypt. Inf. J. (2017).
[40] B. Hamdan, K. Mokhtar, A self-immune to 3D masks attacks face recognition system, Signal Image Video Process. (2018) 1–8.
[41] Y. Wang, S. Chen, W. Li, D. Huang, Y. Wang, Face anti-spoofing to 3D masks by combining texture and geometry features, in: Chinese Conference on Biometric Recognition, Springer, 2018, pp. 399–408.
[42] S. Soltanpour, B. Boufama, Q.J. Wu, A survey of local feature methods for 3D face recognition, Pattern Recognit. 72 (2017) 391–406.
[43] D. Menotti, G. Chiachia, A. Pinto, W.R. Schwartz, H. Pedrini, A.X. Falcao, A. Rocha, Deep representations for iris, face, and fingerprint spoofing detection, IEEE Trans. Inf. ForensicsSecur. 10 (4) (2015) 864–879.
[44] L. Feng, L.-M. Po, Y. Li, X. Xu, F. Yuan, T.C.-H. Cheung, K.-W. Cheung, Integration of image quality and motion cues for face anti-spoofing: a neural network approach, J. Vis. Commun. Image Represent. 38 (2016) 451–460.
[45] O. Lucena, A. Junior, V. Moia, R. Souza, E. Valle, R. Lotufo, Transfer learning using convolutional neural networks for face anti-spoofing, in: International Conference Image Analysis and Recognition, Springer, 2017, pp. 27–34.
[46] R. Shao, X. Lan, P.C. Yuen, Deep convolutional dynamic texture learning with adaptive channel-discriminability for 3d mask face anti-spoofing, in: Biometrics (IJCB), 2017 IEEE International Joint Conference on, IEEE, 2017, pp. 748–755.
[47] R. Shao, X. Lan, P.C. Yuen, Joint discriminative learning of deep dynamic textures for 3D mask face anti-spoofing, IEEE Trans. Inf. ForensicsSecur. (2018).
[48] J. Liu, A. Kumar, Detecting presentation attacks from 3D face masks under multispectral imaging, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2018, pp. 47–52.
[49] S. Liu, P.C. Yuen, S. Zhang, G. Zhao, 3D mask face anti-spoofing with remote photoplethysmography, in: European Conference on Computer Vision, Springer, 2016, pp. 85–100.
[50] X. Li, J. Komulainen, G. Zhao, P.C. Yuen, M. PietikÃn, Generalized face anti-spoofing by detecting pulse from face videos, in: International Conference on Pattern Recognition, 2017, pp. 4244–4249.
[51] S.-Q. Liu, X. Lan, P.C. Yuen, Remote photoplethysmography correspondence feature for 3d mask face presentation attack detection, 2018, pp. 558–573.
[52] J. Hernandez-Ortega, J. Fierrez, A. Morales, P. Tome, Time analysis of pulse-based face anti-spoofing in visible and NIR, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2018, pp. 544–552.
[53] A. Ali, N. Alsufyani, S. Hoque, F. Deravi, Biometric counter-spoofing for mobile devices using gaze information, in: International Conference on Pattern Recognition and Machine Intelligence, Springer, 2017, pp. 11–18.
[54] N. Alsufyani, A. Ali, S. Hoque, F. Deravi, Biometrie presentation attack detection using gaze alignment, in: Identity, Security, and Behavior Analysis (ISBA), 2018 IEEE 4th International Conference on, IEEE, 2018, pp. 1–8.
[55] A. Ali, S. Hoque, F. Deravi, Gaze stability for liveness detection, Pattern Anal. Appl. 21 (2) (2018) 437–449.

[56] N. Alsufyani, A. Ali, S. Hoque, F. Deravi, Biometrie presentation attack detection using gaze alignment, in: Identity, Security, and Behavior Analysis (ISBA), 2018 IEEE 4th International Conference on, IEEE, 2018, pp. 1–8.
[57] D.E. King, Dlib-ml: a machine learning toolkit, J. Mach. Learn. Res. 10 (Jul) (2009) 1755–1758.
[58] I. Chingovska, N. Erdogmus, A. Anjos, S. Marcel, Face recognition systems under spoofing attacks, in: Face Recognition Across the Imaging Spectrum, Springer, 2016, pp. 165–194.
[59] A. Anjos, S. Marcel, Counter-measures to photo attacks in face recognition: a public database and a baseline, in: Biometrics (IJCB), 2011 International Joint Conference on, IEEE, 2011, pp. 1–7.
[60] D. Wen, H. Han, A.K. Jain, Face spoof detection with image distortion analysis, IEEE Trans. Inf. ForensicsSecur. 10 (4) (2015) 746–761.
[61] Z. Boulkenafet, J. Komulainen, A. Hadid, Face anti-spoofing based on color texture analysis, in: Image Processing (ICIP), 2015 IEEE International Conference on, IEEE, 2015, pp. 2636–2640.
[62] L. Mei, D. Yang, Z. Feng, J. Lai, WLD-TOP based algorithm against face spoofing attacks, in: Chinese Conference on Biometric Recognition, Springer, 2015, pp. 135–142.
[63] I. Kim, J. Ahn, D. Kim, Face spoofing detection with highlight removal effect and distortions, in: Systems, Man, and Cybernetics (SMC), 2016 IEEE International Conference on, IEEE, 2016, pp. 004299–004304.
[64] Z. Boulkenafet, J. Komulainen, Z. Akhtar, A. Benlamoudi, D. Samai, S. Bekhouche, A. Ouafi, F. Dornaika, A. Taleb-Ahmed, L. Qin, et al., A competition on generalized software-based face presentation attack detection in mobile scenarios, IJCB 7 (2017).
[65] Z. Boulkenafet, J. Komulainen, A. Hadid, Face antispoofing using speeded-up robust features and fisher vector encoding, IEEE Signal Process. Lett. 24 (2) (2017) 141–145.
[66] F. Peng, L. Qin, M. Long, Face presentation attack detection using guided scale texture, Multimed. Tools Appl. (2017) 1–27.
[67] X. Tu, Y. Fang, Ultra-deep neural network for face anti-spoofing, in: International Conference on Neural Information Processing, Springer, 2017, pp. 686–695.
[68] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, in: International Conference on Learning Representations, 2015.
[69] J. Määttä, A. Hadid, M. Pietikäinen, Face spoofing detection from single images using micro-texture analysis, in: Biometrics (IJCB), 2011 International Joint Conference on, IEEE, 2011, pp. 1–7.
[70] X. Song, X. Zhao, L. Fang, T. Lin, Discriminative representation combinations for accurate face spoofing detection, Pattern Recognit. 85 (2019) 220–231.
[71] M. Sajjad, S. Khan, T. Hussain, K. Muhammad, A.K. Sangaiah, A. Castiglione, C. Esposito, S.W. Baik, CNN-based anti-spoofing two-tier multi-factor authentication system, Pattern Recognit. Lett. (2018).
[72] Y.N. Singh, S.K. Singh, P. Gupta, Fusion of electrocardiogram with unobtrusive biometrics: an efficient individual authentication system, Pattern Recognit. Lett. 33 (14) (2012) 1932–1941.
[73] M. Gomez-Barrero, J. Galbally, J. Fierrez, Efficient software attack to multimodal biometric systems and its application to face and iris fusion, Pattern Recognit. Lett. 36 (2014) 243–253.
[74] P. Wild, P. Radu, L. Chen, J. Ferryman, Robust multimodal face and fingerprint fusion in the presence of spoofing attacks, Pattern Recognit. 50 (2016) 17–25.

Shan Jia received the B.S. degree in electronic and information engineering in 2014 from Wuhan University, Wuhan, China. She is pursuing the Ph.D. degree in the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing (LIESMARS) at Wuhan University, and is currently a visiting student at West Virginia Uni- versity (WVU). Her research interests include face recognition, biometrics, information security, and multimedia processing.



Guodong Guo (M'07-SM'07) received the B.E. degree in automation from Tsinghua University, Bei- jing, China, the Ph.D. degree in pattern recognition and intelligent control from Chinese Academy of Sciences, Beijing, China, and the Ph.D. degree in computer science from University of Wisconsin- Madison, Madison, WI, USA. He is an Associate Professor with the Department of Computer Sci- ence and Electrical Engineering, WVU, Morgan- town, WV, USA. In the past, he visited and worked in INRIA, Sophia Antipolis, France; Ritsumeikan University, Kyoto, Japan; and Microsoft Research, Beijing, China; and North Carolina Central University. He authored a book, Face, Expression, and Iris Recognition Using Learning-based Approaches (2008), co-edited two books, Mobile Biometrics (2017), and Support Vector Machines Applications (2014), and published about 100 technical papers. His research interests include computer vision, machine learning, and multimedia. He received the Outstanding Researcher (2017-2018, 2013-2014), and New Researcher of the Year (2010-2011) at CEMR, WVU. He was selected the "People's Hero of the Week" by BSJB under Minority Media and Telecommunications Council (MMTC) on July 29, 2013. Two of his papers were selected as "The Best of FG'13" and "The Best of FG'15", respectively.

Zhengquan Xu received the B.S. degree in radio technology and information system in 1985, the M.S. degree in communication and electronic system in 1988, both from Tsinghua University, Beijing, China, and the Ph.D. degree in biomedicine engi- neering from Hong Kong Polytechnic University, Hong Kong, China, in 1998. He is currently a professor with the LIESMARS at Wuhan University. His current research interests include information security, cloud computing security, biometrics and multimedia processing.