A 3D Mask Face Anti-spoofing Database with Real World Variations

Siqi Liu[1], Baoyao Yang[1], Pong C. Yuen[1], Guoying Zhao[2]

[1]Department of Computer Science, Hong Kong Baptist University
[2]Department of Computer Science and Engineering, University of Oulu
{siqiliu, byyang, pcyuen}@comp.hkbu.edu.hk, gyzhao@ee.oulu.fi

Abstract

*3D mask face spoofing attack becomes new challenge and attracts more research interests in recent years. However, due to the deficiency number and limited variations of database, there are few methods be proposed to aim on it. Meanwhile, most of existing databases only concentrate on the anti-spoofing of different kinds of attacks and ignore the environmental changes in real world applications. In this paper, we build a new 3D mask anti-spoofing database with more variations to simulate the real world scenario. The proposed database contains 12 masks from two companies with different appearance quality. 7 cameras from the stationary and mobile devices and 6 lighting settings that cover typical illumination conditions are also included. Therefore, each subject contains 42 (7 cameras * 6 lightings) genuine and 42 mask sequences and the total size is 1008 videos. Through the benchmark experiments, directions of the future study are pointed out. We plan to release the database as an platform to evaluate methods under different variations.*

1. Introduction

As the increasing deployment of face recognition in a variety of applications, its security concern becomes increasingly important. Like the other biometric modalities, a major security issue is to detect the spoofing attack. Traditionally, photos and videos are the two medium to carry out the face spoofing attack. In order to detect them, numbers of methods have been proposed and achieved promising results. Nowadays, with the development of 3D reconstruction and 3D print technologies, 3D masks can easily be made with one or few client's face images. Due to the popularity of social networks, those face images can easily be obtained from the Internet. Recently, the affordable off-the-shelf mask was proven to be able to spoof the face recognition system [4]. As a results, a new face spoofing attack—the 3D mask attack has came into our view.

Although the 3D mask spoofing attack becomes the new challenge and attracts more interests in recent years, there are still few methods been proposed to address it. One of the major problem is the lack of sufficient and comprehensive databases due to the expensive price of making a customized 3D mask. Erdogmus *et al*. propose the 3DMAD dataset [4] in the size of 17 subjects with their corresponding masks. The reported LBP based methods achieve encouraging results on it. Though the database size is plentiful considering the high cost of 3D facial masks, the variation of mask type is small. Only ThatsMyFace mask is used which limits the appearance variations.

In the mean time, videos are collected with single camera under studio environment with sufficient soft lighting condition. Actually, in the real application scenario, attackers may spoof the face recognitions system with multi types of 3D mask under different environment. Also, the system may be deployed in different applications with varieties of camera types. Databases contain these variations need to be created as a platform to evaluate the performance of face anti-spoofing methods in real world applications.

However, to the best of our knowledge, we find that existing databases mainly focus on the used fake face while have little concern about the variations in the real world application scenarios, such as the lighting environment and the imaging devices of the face recognition system. For example, most of the existing databases contain limited variations on imaging camera types. The NUAA database [9] as well as the Idiap-Attack Database (Idiap) database use only one web-camera to record the videos. The 3DMAD database captures the color and depth information but also limits to single camera type. For the CASIA database [15], although three cameras are used to represent the low, middle and high imaging quality, the variations are limited on stationary devices. Today, increasing numbers of face recognition related applications are deployed on mobile devices such as smart phone and tablet. The MSU Mobile Face Spoofing Database (MFSD) [13] is build to evaluate the photo and video attacks on mobile devices. However, only two types

of camera are used which is not sufficient enough. Also, current studies ignore the importance of lighting conditions and the databases are collected under small illumination variations. For instance, the NUAA [9], CASIA [15] and MFSD databases are recorded in the same lab light environment. The 3DMAD database is built under the sufficient studio lighting condition. Though the Idiap database [1] mentions the lighting issue which includes the lamp-light and day-light, the variations are still limited.

Consequently, some researches combine the existing databases and use the cross-database testing protocol to simulate the real world scenario. However, for 3D mask attack problem, this limitation is rather severe since there is only one public available database—3DMAD, which is recorded with single camera under fixed studio lighting condition.

Based on the above considerations, we create a new 3D mask face anti-spoofing dataset with more variations as the platform to evaluate the face anti-spoofing methods under more realistic application scenarios. The new database contains 12 masks including two mask types. Meanwhile, we introduce 7 camera types and 6 typical lighting conditions to simulate the real world scenarios. As a result, each subjects contains 42 genuine and attack videos with different variations. Two new testing protocols are carefully designed to evaluate the performance of face anti-spoofing methods when confront numbers of variations. We further analyze the effectiveness of LBP based methods as a benchmark to the proposed dataset.

The organization of this paper is as follows. We review the related databases and methods in Section 2. In Section 3 we analyze our new database through the three variations mentioned above. In Section 4, we introduce the test protocols . In Section 5 we report the benchmark experiments. Finally, we conclude this paper by drawing few remarks in Section 6 .

2. Related works

3D mask face anti-spoofing becomes increasingly important as the ThatsMyFace mask has been proved to be able to spoof the face recognition system. As a new research topic, only one database, the 3DMAD [4], is built to address this problem. The 3DMAD contains 17 subjects and each of them corresponds to a customized 3D facial mask from ThatsMyFace company. Considering the expensive price of masks, the database is plentiful in size although it is relatively small compare with the existing prints or video based databases. The database is recorded with Microsoft Kinect under the same well optimized studio lighting condition. Existing face anti-spoofing databases mainly concentrate on the print and video attack. The NUAA database [9] contains over 50K images from 15 subjects. Although the size is large, the database is build on images and it is not public available. The Idiap database [1] contains 50 subjects and 400 videos in total. The CAISA database [15] combines three types of attack together to simulate the application scenario. Recently, due to the increasing popularity on mobile face recognition, the MFSD database [13] is created for the analysis of face anti-spoofing about mobile devices. Two types of cameras are used including the built-in camera in Laptop and front-facing camera in mobile phone.

While most of existing methods mainly focus on how to detect different types of spoofing attack, they pay little attention on the affect of real-world environmental variations. Although the CASIA database [15] considers the variation of imaging quality and introduces three cameras to cover the low, middle and high quality levels; the Idiap database [1] introduces two typical in-door lighting conditions; and the MFSD database employs two cameras, their variations are still limited.

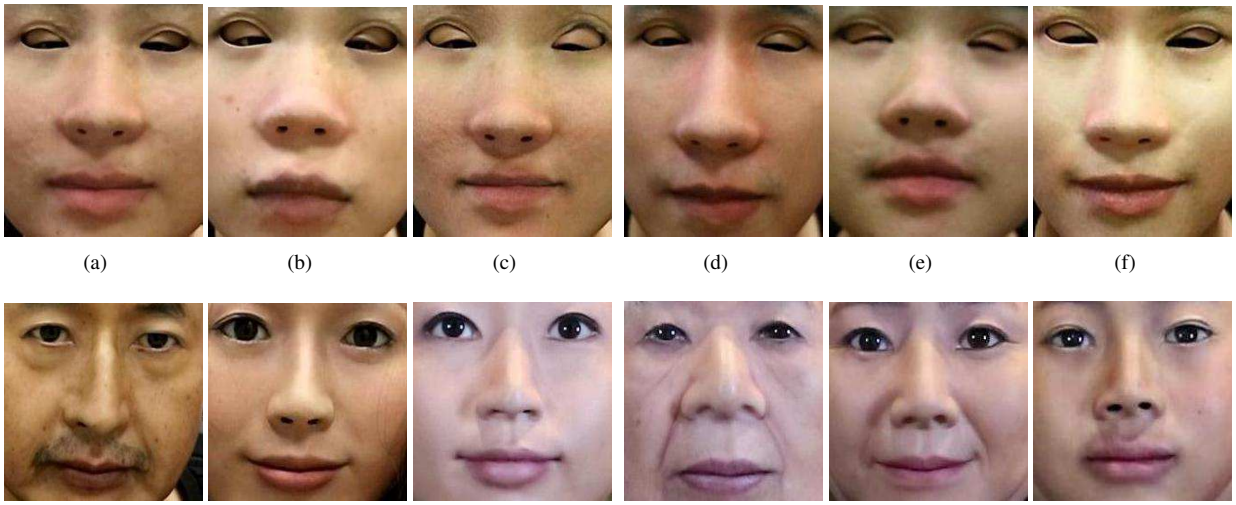
As a new research topic, there are few methods be proposed to target 3D mask face anti-spoofing. Existing methods can be mainly divided into two categories: texture based approaches and motion based approaches. The texture based methods use the appearance detail differences to detect printed or displayed attacks. The Multi-Scale LBP [6] concatenates different LBP settings to enhance the effectiveness and achieves encouraging results on 3DMAD. Deep learning based method [7] also achieves promising results recently. While in this paper, we will not evaluate it due to the the database size. Besides, the image distortion analysis (IDA) based methods achieve good results on MSFD [13]. But for the 3D mask attack, they may not perform well since the 3D mask has little relation to the image quality.

Motion-based approaches use human-computer interaction (HCI) or unconscious face motion to detect photo and video attacks [8, 3, 5]. These approaches are particularly effective against photo and stationary screen attacks. However, this may not stand if we simply cut the eyes or mouth part out on the spoof medium. Also, the motion related 3D structure recover based methods [12] may not be effective since the 3D masks move in the similar way compare to real faces.

For other methods, spectrum analysis [14] seems feasible on detecting the 3D mask attack, but it is not convenient to be applied in existing color image based face recognition systems.

3. The New Database

In this section, we introduce our new 3D mask face anti-spoofing database. We first describe the two types of 3D mask. Then, we introduce the variation of camera under application scenarios. Finally, we analyze the employed lighting variation. The recording setting details are given at last.



(a) (b) (c) (d) (e) (f)

(g) (h) (i) (j) (k) (l)

Figure 1. Sample mask images in the proposed new 3D mask face anti-spoofing database. (a)-(f) are ThatsMyFace masks and (g)-(l) are Real-F masks.

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

Figure 2. High resolution sample images of Thatsmyface (TF) mask (a) and REAL-F mask (b).

3.1. Mask types

Under real world scenarios, the face recognition system may confront different types of 3D mask attacks. In the proposed database, we increase the diversity of mask types by employing masks with different quality from two companies. One is the ThatsMyFace which is convenient to be obtained from user perspective and the other is the REAL-F with exquisite appearance quality.

3.1.1 The ThatsMyFace Mask

The ThatsMyFace mask uses the 3D reconstruction and 3D printing techniques to sculpt the customized facial mask. Unlike the mask technique mentioned in [1], users only need to upload one front face image and add few annotations to locate the facial structure and boundaries. Then, the 3D facial mask model will be generated. If users are satisfied with the model, it will be implemented with 3D printer. The advantage of ThatsMyFace is the convenience of making a mask with single image. However, due to the defects of 3D printing technique, we can find its skin texture difference, which results in the promising performance of LBP based methods reported in [4]. Fig. 2(a) presents the sample image of ThatsMyFace mask.

3.1.2 The REAL-F Mask

In order to enlarge the mask type diversity, we introduce the Real-F 3D facial mask[1] in our new dataset. The Real-F mask has exquisite appearance quality which looks very similar as the genuine face. This precise appearance duplication is based on the Three-Dimension Photo Form (3DPF) technique. It can transfer the 2D facial image onto the 3D prototype which makes it possible to reproduce the skin texture and even the eyes' blood vessels and iris. Fig 2(b) demonstrates the appearance quality of the Real-F mask. For the construction of 3D prototype, users need to take a 3D scan or facial impression. Although the 3D scan requires users' cooperation which is hard to achieve from the attacker's perspective as mentioned in [4], this step could be improved and replaced by advanced 3D reconstruction techniques. Our focus here is to evaluate the face anti-spoofing methods with different types of masks.

In the proposed database, we have 6 different masks for the two mask types corresponding to the 12 subjects.

3.2. Cameras

The imaging devices decide the image or video quality of the face anti-spoofing system. As mentioned in [15], the performance of a method, especially for the texture based

[1] real-f.jp\en_the-realface

one, depends on the image quality to some extends. In the application scenario, face recognition system may be deployed with numbers of devices such as the personal computer (PC), public face recognition system such as immigration, or even portable digital camera. Thus, we introduce 3 devices to cover these typical scenarios. A web camera Logitech C920 is used to represent the personal use face recognition scenario. The video is recorded in the size of 1280*720. For the public system, we choose an economic class industrial camera and record in 800*600 resolution. For portable digital camera, we select the Canon EOS M3, a new mirrorless camera and record under 1280*720.

Recently, as the development of mobile device, increasing numbers of applications with face recognition function are designed. Face anti-spoofing on mobile device therefore attracts more interests. In the new database, we carefully select 4 mobile devices to simulate these scenarios. They are three smart phones: Nexus 5, iPhone 6, Samsung S7, and one tablet: Sony Tablet S. All of them are in default settings when recording videos.

The sample images (only face region) of the 7 devices are presented in Fig. 3

(a) Logitech C920 (b) Industrial Cam. (c) EOS M3

(d) Nexus 5 (e) iPhone 6 (f) Samsung S7 (g) Sony Tablet S

Figure 3. Sample face images recorded by different cameras under same lighting condition (room light).

3.3. Lighting conditions

The Lighting condition is another variation that rarely be mentioned in face anti-spoofing databases. Actually, the performance of existing methods may vary when the lighting changes. Under a sharp side light, the facial structure may cast shadow on the skin (see Fig. 4(e)) and affect the results. Furthermore, with different illumination, the imaging quality may vary since most of the devices have their own algorithms to adjust the exposure level. For instance, as shown in Fig. 4(b) and 4(c), the texture details are blurred in low light environment since the built-in exposure compensation algorithm.

Considering the application environment, we design 6 illumination conditions to cover the typical scenes when use the system on whether stationary or mobile devices. Sample images are shown in Fig. 4. Room light is the most common condition we use the system, we set the scene directly using the office light. Low light, bright light and warm light are the typical variations of the in-door light. Besides, we introduce the side light and up side light to simulate the possible harsh lighting condition, e.g., lamp light in the dark room and the direct out-door sun light. All the five lighting variation are set up with the adjustable camera lights.

(a) Room light (b) Low light (c) Bright light

(d) Warm light (e) side light (f) Up side light

Figure 4. Sample face images recorded under different lighting conditions. Images are captured by Sony Tablet S.

3.4. Recording settings

The proposed 3D mask face anti-spoofing database contains 12 subjects with 12 masks from 2 companies. Half are from ThatsMyFace and half are from Real-F. The sample images captured by EOS M3 are shown in Fig. For each subject, we record 42 videos as the combination of 7 cameras and 6 lighting variations. As a result, the database contains 1008 videos in total. The time length of each video is around 10 seconds. For the frame speed, except the EOS M3 is in 50 fps and the industry camera is in 20 fps, all other devices are set in 30 fps. The resolution settings are described in Section 3.2.

When recording the videos, we use a tripod to fix the stationary devices: web camera Logitech C920, industrial camera and digital camera Canon EOS M3. We let the subject sit around 80 cm away from the camera. For mobile devices: Nexus 5, iPhone 6, Samsung S7 and Sony Tablet S, we ask the subject to hold it buy hand and stay in a comfortable face pose. This results in a little pose variations since the camera position is usually lower for mobile devices compare with the fixed stationary one.

For the set up of lighting strength such as low light and bright light, we use a photometer to do the measurement. For room light the strength on face is around 250-300 lux, so we set the low light around 150 lux with the camera light. We set the bright light on face for round 1000 lux. For bright light, low light and warm light, two camera lights are used and set in an appropriate position to minimize the shadow cast on face. For side light, we use only one to shoot from a side, parallel to the face. For up side light, two lights are lifted to 2 meter height through tripod and shoot from the top.

4. Testing Protocols

In this section, we design two testing protocols from different perspective. One is the intra variation testing to evaluate algorithms under constrained environment. The other is the cross variation testing to evaluate the performance of methods when confronting real world scenarios.

4.1. Intra-Variation Test

For intra variation experiment, we only focus on one single variation of a variation type in each sub-experiment. This is for the evaluation of face anti-spoofing methods under controlled conditions. Thus, considering the numbers of three variation types, there are 2*7*6 intra variation sub-experiments since we have 2 mask types, 7 cameras and 6 lighting conditions respectively. For instance, one of the sub-experiments is to do the evaluation for web camera confronts ThatsMyFace mask attack under room lighting condition. In each sub-experiment, we basically follow the LOOCV protocol defined in [4] using half subjects as training set and rests are development sets after selecting one as the testing subject. Specifically, in order to avoid the affect of subjects' sequence, we randomly divide the subjects into training and development sets after leaving one out for testing. We take the average 100 random LOOCV to achieve the final results.

4.2. Cross-Variation Test

For the cross variation evaluation, we design a new protocol call leave one variation out (LOVO). In each iteration, we choose one variation from one variation type as the training set and use the rest (in the same variation type) as the testing set. For other variations, we only consider one of them in each sub-experiment. For example, when do the LOVO on camera types, there will be 2*6 (mask types * lightings) sets of results and for the LOVO on mask types we have 6*7 (lightings * cameras) sets of results. For the LOVO on cameras and lighting variations, we use all subjects corresponding to each type of mask as the training or testing sets. The cross variation test is to simulate the application scenarios since in the real world applications, we may confront numbers of variations while only have limited samples under specific constrains to train the model.

5. Experiments

5.1. Baseline methods

Based on the analysis in Section 2, we choose the texture based methods as the baseline. In particular, the Multi-Scale LBP (msLBP) [6], transitional LBP (tLBP) and modified LBP (mLBP) [10], are selected as they achieve promising results on the 3DMAD database [4]. For msLBP, we follow the same setting in [6] which concatenated the 3×3 $LBP^{u2}_{8,1}$, $LBP^{u2}_{8,2}$ and $LBP^{u2}_{16,2}$ as an enhanced 833-dimension feature. For tLBP and mLBP, we divide the image into 3×3 blocks and calculate the feature on each of them. The final representation is also formed by concatenation. Uniform mapping is also applied on them. For the input image, we manually select one clear (without motion blur) frame from each video and use Viola-Jones face detector [11] to extract face image. Then, we follow [4] to crop face image to focus on the evaluation of mask appearance. Result input image samples can be view in Fig. 1.

For the classification, we adopt the support vector machine (SVM) with RBF kernel.

5.2. Results & Analysis

In this section, we report the baseline and experiments based on the testing protocols designed in Section. Moreover, we add and overall test [2] and the average of intra testing to compare the performance between the baseline methods. Note that the intra testing results are the average of all intra sub-experiments and the over testing results are counting all variations simultaneously when applying LOOCV.

The experimental results of msLBP under the above mentioned two testing protocols are summarized in Table 1 and Table 2, 3, 4, respectively. We use L1-L6 to represent different lighting condition: room light, low light, bright light, warm light, side light, and up side light, respectively. For the intra variation experiments, we find that the EER of ThatsMyFace attack is lower than the Real-F one. As mentioned before, the appearance details of Real-F mask is more exquisite compared with the ThatsMyFace mask so that the performance of texture based methods may drop as the increasing of appearance quality. To compare the EER along different camera types, we find that the performance is quite similar. It may because of the adapt ability of devices under different lighting conditions. In other words, cameras will preserve similar detail information and results in the close performance. Also, we find that the performance changes are different among cameras under varies of lighting conditions. It may indicate the difference of exposure adjustment algorithms for different cameras. Specially, the industry camera is set in MF mode with fixed exposure speed and compensation level while results are also similar. This indicates the discriminative ability of msLBP on intra variation scenario. For the cross variation experiments, we

	ThatsMyFace						Real-F					
	L1	L2	L3	L4	L5	L6	L1	L2	L3	L4	L5	L6
WebCam	30.6	33.1	33.1	33.5	30.7	22.2	32.6	31.9	31.4	31.7	24.4	24.6
IndCam	30.4	32.3	34.8	30.3	13.8	30.8	37.0	40.8	40.3	38.2	36.4	48.3
Canon	23.8	31.7	29.0	34.5	32.7	41.1	27.4	39.1	31.1	35.8	27.9	38.4
iPhone	32.9	15.5	22.8	29.8	37.7	34.6	55.7	22.8	35.5	41.3	22.3	44.1
Nexus	26.2	22.0	39.5	18.4	30.9	34.6	35.7	37.0	42.9	67.6	33.3	47.4
Samsung	26.2	22.0	39.5	18.4	30.9	34.6	42.1	20.7	39.6	64.5	20.3	42.4
Sony	36.8	38.3	36.5	23.8	32.5	34.6	45.1	47.4	53.5	41.4	38.7	55.8

Table 1. The EER(%) of intra-variation experiments on two types of mask attack

	WebCam	IndCam	Canon	iPhone	Nexus	Samsung	Sony
ThatsMyFace	20.0	40.0	30.6	27.9	34.0	32.8	32.4
Real-F	34.3	36.2	36.2	38.6	38.1	36.7	36.7

Table 2. The EER(%) of cross lighting experiment

	L1	L2	L3	L4	L5	L6
ThatsMyFace	32.9	38.3	42.5	33.9	35.1	38.9
Real-F	32.7	41.5	36.3	34.3	32.1	43.9

Table 3. The EER(%) of cross camera experiment

	L1	L2	L3	L4	L5	L6
WebCam	41.7	50.0	50.0	25.0	41.7	50.0
IndCam	50.0	50.0	50.0	50.0	50.0	50.0
Canon	41.7	50.0	50.0	50.0	50.0	41.7
iPhone	41.7	25.0	50.0	50.0	33.3	50.0
Nexus	33.3	16.7	50.0	50.0	41.7	41.7
Samsung	41.7	41.7	41.7	50.0	50.0	58.3
Sony	29.2	50.0	50.0	50.0	41.7	50.0

Table 4. The EER(%) of cross mask type experiment

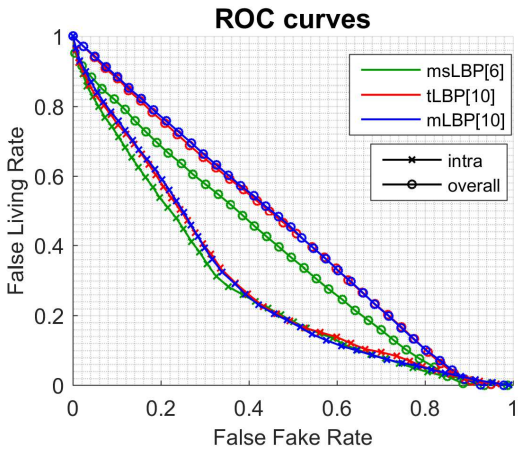


Figure 5. ROC curves of three baseline methods under intra and overall testing protocol

can obviously find the performance drops through the cross mask experiment. For almost all variation combinations, the EER is around 50%. This implies that the texture based methods can not generalize well while confronting different mask appearance. Also, it is easy to identify the 30% to 40% performance decrease through the results of the cross camera and cross lighting condition listed in Table 2 and 3. Note that the training and testing samples under these protocols contain the same face expect the image quality. Thus, the degradation is entirely caused by the image quality difference among cameras and the lighting variations which shows the limitation of generalize ability again.

The Fig. 5 compares the ROC curves of three LBP based methods under overall and average of intra testing protocol.

6. Conclusion

In this paper, we build a new 3D mask face anti-spoofing database with more variations to simulate the real world scenario. Our database contains 2 types of mask in different appearance quality, 7 cameras from the stationary and mobile devices and 6 lighting settings that cover typical illumination conditions. Texture based methods are selected as the benchmark in experiments. Results show their effectiveness under intra variation testing and also expose the weakness under real world situations. A practical direction is to enhance the generalize ability under numbers of variations. The proposed database is planed to be released as a platform for the evaluation of 3D mask face anti-spoofing under real world scenario.

7. Acknowledgement

This project is partially supported by Hong Kong RGC General Research Fund HKBU 12201215.

References

[1] I. Chingovska, A. Anjos, and S. Marcel. On the effectiveness of local binary patterns in face anti-spoofing. In *Biometrics Special Interest Group (BIOSIG), 2012 BIOSIG-Proceedings of the International Conference of the*, pages 1–7. IEEE, 2012. 2, 3

[2] T. de Freitas Pereira, A. Anjos, J. M. De Martino, and S. Marcel. Can face anti-spoofing countermeasures work in a real world scenario? In *Biometrics (ICB), 2013 International Conference on*, pages 1–8. IEEE, 2013. 5

[3] T. de Freitas Pereira, J. Komulainen, A. Anjos, J. M. De Martino, A. Hadid, M. Pietikäinen, and S. Marcel. Face liveness detection using dynamic texture. *EURASIP Journal on Image and Video Processing*, 2014(1):1–15, 2014. 2

[4] N. Erdogmus and S. Marcel. Spoofing face recognition with 3d masks. *Information Forensics and Security, IEEE Transactions on*, 9(7):1084–1097, 2014. 1, 2, 3, 5

[5] K. Kollreider, H. Fronthaler, M. I. Faraj, and J. Bigun. Real-time face detection and motion analysis with application in liveness assessment. *Information Forensics and Security, IEEE Transactions on*, 2(3):548–558, 2007. 2

[6] J. Määttä, A. Hadid, and M. Pietikainen. Face spoofing detection from single images using micro-texture analysis. In *Biometrics (IJCB), 2011 international joint conference on*, pages 1–7. IEEE, 2011. 2, 5

[7] D. Menotti, G. Chiachia, A. Pinto, W. Robson Schwartz, H. Pedrini, A. Xavier Falcao, and A. Rocha. Deep representations for iris, face, and fingerprint spoofing detection. *Information Forensics and Security, IEEE Transactions on*, 10(4):864–879, 2015. 2

[8] G. Pan, L. Sun, Z. Wu, and S. Lao. Eyeblink-based anti-spoofing in face recognition from a generic webcamera. In *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, pages 1–8. IEEE, 2007. 2

[9] X. Tan, Y. Li, J. Liu, and L. Jiang. Face liveness detection from a single image with sparse low rank bilinear discriminative model. In *Computer Vision–ECCV 2010*, pages 504–517. Springer, 2010. 1, 2

[10] J. Trefnỳ and J. Matas. Extended set of local binary patterns for rapid object detection. In *Proceedings of the computer vision winter workshop*, volume 2010, 2010. 5

[11] P. Viola and M. J. Jones. Robust real-time face detection. *International journal of computer vision*, 57(2):137–154, 2004. 5

[12] T. Wang, J. Yang, Z. Lei, S. Liao, and S. Z. Li. Face liveness detection using 3d structure recovered from a single camera. In *Biometrics (ICB), 2013 International Conference on*, pages 1–6. IEEE, 2013. 2

[13] D. Wen, H. Han, and A. K. Jain. Face spoof detection with image distortion analysis. *Information Forensics and Security, IEEE Transactions on*, 10(4):746–761, 2015. 1, 2

[14] D. Yi, Z. Lei, Z. Zhang, and S. Z. Li. Face anti-spoofing: Multi-spectral approach. In *Handbook of Biometric Anti-Spoofing*, pages 83–102. Springer, 2014. 2

[15] Z. Zhang, J. Yan, S. Liu, Z. Lei, D. Yi, and S. Z. Li. A face antispoofing database with diverse attacks. In *Biometrics (ICB), 2012 5th IAPR international conference on*, pages 26–31. IEEE, 2012. 1, 2, 3