3DMA: A Multi-modality 3D Mask Face Anti-spoofing Database

Jinchuan Xiao [1,2], Yinhang Tang[3], Jianzhu Guo[1,2], Yang Yang [1], Xiangyu Zhu[1], Zhen Lei[1], Stan Z. Li[1]
[1]National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences
[2]University of Chinese Academy of Sciences
[3]AuthenMetric Inc.
{jinchuan.xiao, jianzhu.guo, yang.yang, xiangyu.zhu, zlei, szli}@nlpr.ia.ac.cn

Abstract

Benefiting from publicly available databases, face anti-spoofing has recently gained extensive attention in the academic community. However, most of the existing databases focus on the 2D object attacks, including photo and video attacks. The only two public 3D mask face anti-spoofing database are very small. In this paper, we release a multi-modality 3D mask face anti-spoofing database named 3DMA, which contains 920 videos of 67 genuine subjects wearing 48 kinds of 3D masks, captured in visual (VIS) and near-infrared (NIR) modalities. To simulate the real world scenarios, two illumination and four capturing distance settings are deployed during the collection process. To the best of our knowledge, the proposed database is currently the most extensive public database for 3D mask face anti-spoofing. Furthermore, we build three protocols for performance evaluation under different illumination conditions and distances. Experimental results with Convolutional Neural Network (CNN) and LBP-based methods reveal that our proposed 3DMA is indeed a challenge for face anti-spoofing. This database is available at http://www.cbsr.ia.ac.cn/english/3DMA.html. We hope our public 3DMA database can help to pave the way for further research on 3D mask face anti-spoofing.

1. Introduction

Recently, the face recognition [9, 19, 12, 6] research develops rapidly and has been an extremely reliable technology in a variety of challenging applications. However, the various presentation attacks threaten the existing face recognition systems. Nowadays, the threat from the spoofing attacks, such as face images, videos or 3D masks of legitimate users, has been realized. And the face spoofing detection (also known as presentation attack detection, PAD) has achieved a lot of attentions [23, 1, 7].

978-1-5386-9294-3/18/$31.00 ©2019 IEEE

(a) VIS image of a mask

(b) NIR image of a mask

(c) VIS image of a real subject

(d) NIR image of a real subject

Figure 1: VIS and NIR images of a mask and a real subject.

Considering the diversity of the fake faces attacks, the types of presentation attacks and the fake subjects should become as many as possible in our research of countermeasures. Meanwhile, the anti-spoofing system should be robust to environmental variations, such as lighting conditions, distance variation, etc. However, the public face anti-spoofing databases and the recent work about face anti-spoofing mostly focus on 2D attacks (photos, video playbacks). Due to the high cost of 3D mask, most public databases, such as the CASIA FASD database [23], and the OULU-NPU database [1], have no concern about the problem from the mask attack. Meanwhile, the quantity of the recorded mask subjects in public databases is limited. The limited number of fake subjects limits the performance and application environment of those proposed algorithms. Besides, the anti-spoofing systems which work with the infrared spectrum are usually invariant to illumination changes in the environment, and also are naturally resistant

to many 2D spoofing attacks [3]. However, most existing databases only cover one modality (visual spectrum), there are just a few databases cover multi-spectrums [5].

We address these problems above and collect a new 3D mask face anti-spoofing database covering visual (VIS) spectrum and near-infrared (NIR) spectrum. The database contains 920 video samples collected from 67 genuine subjects and 48 masks. We consider two lighting and four distance settings to evaluate the generalization of different methods. Moreover, we also provide three protocols for evaluation as follows. All the subjects in the collected database are divided into the training set, validation set, and test set. Wherein, the test set contains 18 genuine subjects and 18 3D masks, the validation set includes 17 genuine subjects and 17 3D masks, and the training set includes the others. We mainly use the LBP based method and the end to end CNN-based method to report the baseline performance.

The rest of this paper is organized as follows. Section 2 introduces the related anti-spoofing algorithms and publicly available anti-spoofing databases. Section 3 gives the details of our proposed 3D mask database and describes the testing protocols. We show the baseline experimental results in Section 4. Section 5 concludes the paper.

2. Related Works

2.1. Publicly Available Anti-spoofing Databases

Existing publicly available databases play an important role in finding the best countermeasure for face anti-spoofing. Zhang *et al*. [24] released a face anti-spoofing database named CASIA Face Anti-Spoofing Database. In [1], the authors introduced a new public face anti-spoofing database named OULU-NPU. To simulate a real-world scenario, it considers three variations in the capture conditions including illumination variation, recording devices and presentation attack instruments.

Benefiting from the new sensors in Microsoft Kinect, multi-modality data like visual image, infrared image and depth image can be acquired for face presentation attacks detection. Hence, the anti-spoofing countermeasures related to the multimodalities images had also been addressed in several works. The Msspoof database [3] only includes print attacks with the modalities of VIS videos and NIR videos. The CASIA-SURF database collected in [21] is a large-scale 2D face presentation attacks detection database with three modalities.

Besides the printed attacks and video attacks detection appearing in the three above public databases, 3D mask attacks are equally received more attention. The first public 3D mask attacks database is 3DMAD presented in [5]. Erdogmus *et al*. applied Microsoft Kinect to acquire color and depth modalities samples for creating the 3DMAD database. In [15], Liu *et al*. build a 3D mask presentation attacks detection database, which employs seven cameras to record videos in six illumination conditions for simulating the real-world scenario.

However, most of the existing publicly available databases have only a small number of subjects and videos. The lack of publicly available data is an apparent reason hindering the development of face presentation attacks detection technology.

2.2. Anti-spoofing Methods

The texture is a very discriminating hint for distinguishing between genuine faces and fake faces. Many previous researches used some high-level texture features as a basis for classification, such as LBP [4], HOG [13] and SIFT [17]. Although the texture information is effective for face anti-spoofing, its performance is sensitive to environment variations, including illumination and distance changes.

In [18], authors propose the image quality analysis based methods. It shows that the image quality analysis based methods have good generalization performances for spoofing face detection.

The subtle movements on faces are also a useful cue to classify genuine and fake faces. In [2], the motion information like eye blinking and lips movement is extracted for spoofing face detection. The motion information can be easily imitated by video replay presentation.

There are many recent works [16, 11] apply deep learning based algorithms to resolve the anti-spoofing problem. Initially, most of the works take the anti-spoofing problem as a simple classification task and train the classification network with the softmax loss. Recently, Liu *et al*. [16] indicate the importance of auxiliary supervision. In [11], Jourabloo *et al*. treat the face anti-spoofing as a similar problem to the denoising and deblurring technology. A CNN model is learned to estimate such spoofing noise from a fake sample, and then the estimated spoofing noise is used for classification.

All the above works just utilize the visual spectrum. However, the multi-spectrum based methods can involve more texture information than the visual spectrum, which results in better anti-spoofing performance. Yi *et al*. [20] propose a multi-modalities camera system covering VIS and NIR spectrums to detect printed photo attacks. Zhang *et al*. [24] chose 850 nm and 1450 nm as the wavelengths to be the supplement to VIS. In [14], authors propose a multispectral imaging approach to more accurately detect 3D mask attacks. Ranges of CNN-based configurations are investigated to improve the detection accuracy from such presentation attacks.

From the related works above, we suppose that deep learning based methods and LBP based methods are widely used in face presentation attacks detection methods. So we

Table 1: The comparison of the publicly available anti-spoofing databases (* indicates that Msspoof only contains images).

Database	Year	# of subjects	# of videos	Camera	Modality types	Spoofing attacks
CASIA-MFSD [24]	2012	50	600	VIS	VIS	Print, Replay
Oulu-NPU [1]	2017	55	5940	Phone	VIS	2 Print, 2 Replay
SiW [16]	2018	165	4620	VIS	VIS	2 Print, 4 Replay
Msspoof [3]	2016	21	4704*	uEye camera	VIS/NIR	Print
CASIA-SURF [21]	2018	1000	21000	RealSense	VIS/Depth/NIR	Print, Cut
3DMAD [5]	2013	17	255	Kinect	VIS/Depth	3D Mask
HKBU-MARs [15]	2016	12	1008	VIS	VIS	3D Mask
3DMA(ours)	2019	67 genuine + 48 masks	920	AuthenMetric binocular camera	VIS/NIR	3D Mask

consider face anti-spoofing as a binary classification problem and apply two methods based on deep learning and LBP respectively as the benchmark algorithms.

In this work, we collect a novel database containing multi-spectrum data which focuses on the 3D mask attacks detection. A series of benchmark experiments are performed using LBP-based method and deep learning based method. The contributions of this work include: 1) A novel 3D mask anti-spoofing database containing multi-spectrum data are released. 2) Several experimental protocols related to the proposed database are provided considering the variety of subjects, lighting conditions and capture distance. 3) Benchmark experiments using LBP-based and CNN-based face anti-spoofing algorithms have been conducted as the baseline performance of this database.

3. The Collected Database

3.1. Genuine Subjects and 3D Masks

A total of 67 genuine subjects participate in the database collection, and 48 kinds of 3D face masks are collected. All of the subjects are Chinese people in the ages between 20 and 40 years old. The 3D face masks used have no intersections with these 67 subjects. One subject is asked to wear a 3D mask to perform the face anti-spoofing attack.

3.2. Lighting Condition Settings

Due to the different lighting conditions, the texture details of the subjects imaging might be different. For example, when the light intensity of the environment is strong, the texture details of the acquired samples will be clear. On the contrary, the weak ambient light intensity will blur the texture details. The face presentation attacks detection algorithms should be robust to such light intensity change in real application. Two lighting conditions (i.e., 200 lux and 500 lux) are considered, in which the previous lighting condition simulates indoor environment with normal brightness, and the latter lighting condition simulates bright indoor environment.

3.3. Distance Settings

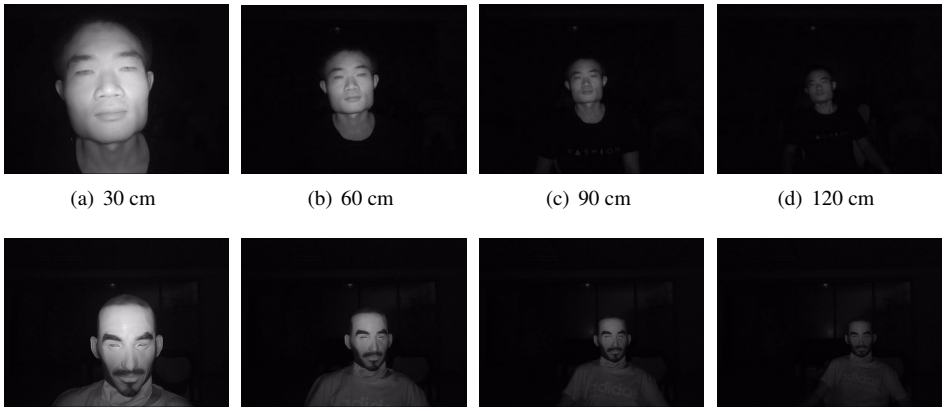
The distance between the camera and the collecting subjects is another environmental factor affecting the imaging quality during acquisition. It also has a significant impact on the face presentation attacks detection algorithm performance. There are mainly two reasons. The first one is that the captured face image size is different at different distances so that the number of effective pixels in the face region is different. The second reason is that the near-infrared imaging relies on active near-infrared light on the camera. Due to the power of the NIR bulb, the NIR light attenuates if the distance increases so that the NIR imaging is affected. Four distance between the camera and the face is considered: 1) 30 cm; 2) 60 cm; 3) 90 cm; 4) 120 cm. Some sample images are shown in Figure 2.

3.4. Recording Settings

We use the R0710A binocular camera which is designed and fabricated by AuthenMetric for acquisition. To ensure that the light intensity meets the imaging requirements, we used the photometer to measure the light intensity before the face region. We collect the database in an indoor environment with curtains. The light intensity is between 180 and 230 lux or 470 and 540 lux. In each lighting condition, four videos of 300 frames with four distance settings are recorded for each person in about ten seconds. During the acquisition, subjects are required to face the camera, and there is no angle deflection. The acquisition image resolution is set to 640×480.

3.5. Data Preprocessing

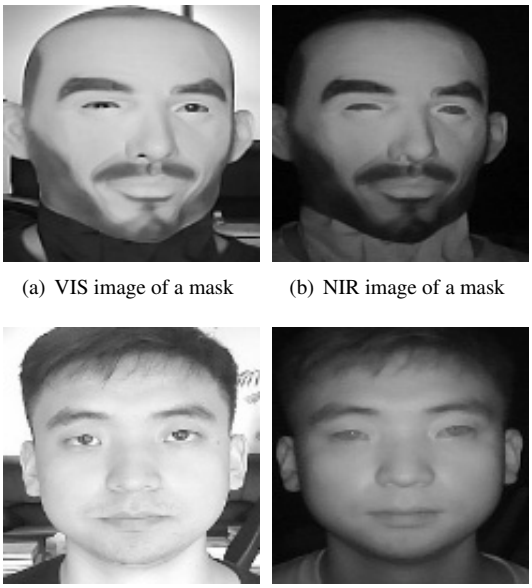
We first use Faceboxes [22] to detect faces on both the VIS image and the NIR image from the same frame in the captured video. Second, we extend the length and the width of face boxes to 1.3 times for ensuring the extracted image samples containing the whole face area. No face alignment operator is adopted. In order to exclude the color bias between the genuine faces and masks, we convert the extracted images achieved above into grayscale images and resize the extracted images to 120×120. The preprocessed



(a) 30 cm (b) 60 cm (c) 90 cm (d) 120 cm

(e) 30 cm (f) 60 cm (g) 90 cm (h) 120 cm

Figure 2: NIR images of real people under different distance conditions.



(a) VIS image of a mask (b) NIR image of a mask

(c) VIS image of a real subject (d) NIR image of a real subject

Figure 3: Examples of cropped face images.

image from each frame in the video samples contains at least one detected face. Some of the images of the face area are shown in Figure 3.

3.6. Evaluation Protocols

All subjects in the database are divided into three disjoint sets, which are used as the test set, validation set, and training set. The test set includes 18 genuine subjects and 18 3D facial masks. The validation set includes 17 genuine subjects and 17 3D facial masks. There is no intersection between the test set and the validation set in terms of subjects and images. The training set includes other subjects. Specifically, all the subjects in the test set and the validation set are recorded under two lighting conditions. Most of the subjects in the training set are recorded under two lighting conditions, while there are a small number of genuine subjects in the training set are recorded under only one lighting condition.

Three test protocols are provided to evaluate the performance of face presentation attacks detection methods from several aspects. The first test protocol is designed to evaluate the generalization capability of face presentation detection methods across different subjects. Therefore, all the videos captured in each lighting condition and distance condition in training, validation and test set are used.

The second test protocol focuses on the effect of lighting variations on the performance of the face presentation attacks detection algorithm. As the samples are recorded with two different illumination settings, we construct the train, validation and evaluation sets using the videos which are recorded under different lighting conditions.

The third protocol is designed to evaluate the generalization capability of the face presentation attacks detection methods across four different distances. Specifically, in each experimental group, we choose the samples captured in one certain distance to build the train set ant the validation set. The samples captured in the other three distances are built as the test set. The settings of the three test protocols presented above are detailed in Table 2.

4. Baseline Algorithms

4.1. LBP-based Methods

LBP has been quite successful in face presentation attacks detection tasks [4], so we use it as the benchmark algorithm for evaluating the performance. Each frame of the video is processed using the LBP operator. Then we calcu-

Table 2: The details about the use of videos in three test protocols.

Protocol	Subset	Subjects	Lighting Conditions (lux)	Distance Conditions (cm)	Real Videos	Mask Videos	Total Videos
Protocol I	Test	1∼18	All	All	144	144	288
	Val	19∼35	All	All	136	136	272
	Train	All the others	All	All	256	104	360
Protocol II	Test	1∼18	200/500	All	72	72	144
	Val	19∼35	500/200	All	68	68	136
	Train	All the others	500/200	All	128	52	180
Protocol III	Test	1∼18	All	(60,90,120)/(30,90,120) /(30,60,120)/(30,60,90)	108	108	216
	Val	19∼35	All	30/60/90/120	34	34	68
	Train	All the others	All	30/60/90/120	64	26	90

late the statistical histogram of the processed image which is used as the 256-dimensional feature for the classification. Finally, we apply SVM with radial basis function (RBF) kernel as the classifier. In the later experiments, we set the penalty factor C to 1, and the tolerance for stopping criterion is set to 1e-3.

4.2. CNN-based Methods

In the baseline experiment, we treat the face anti-spoofing problem as a binary classification problem. Hence we use an end-to-end method based on the ResNet network to detect the 3D mask attack. The network structure used in the experiment is based on ResNet-9 network which consists of two convolution layer, three convolutional blocks, and a global average pooling layer. For evaluating the baseline, we first separately use the NIR image and VIS image as the input to the network. Then we propose three fusion methods as baseline methods to fit the NIR+VIS dual-modalities based anti-spoofing problem.

The first fusion baseline method is the Concatenation Data Fusion method. In this method, we concatenate the corresponding pair of the VIS image and the NIR image as a two-channel input of the classification network. The second fusion baseline method is the Naive Data Fusion method. In this method, we use a shared model to learn the data of the VIS image and the NIR image in one pair. Then the pair of images from two sources are used separately as one single channel input of the shared model. The last fusion baseline method is the Squeeze and Excitation Fusion method which is inspired by [8, 21]. We design a two-stream network architecture. Each sub-network is fed with the image in one modality. Then the Squeeze-and-Excitation branch [8] is applied to adjust the weight of different features from each channel, which is different from [21] because in [21] the Squeeze-and-Excitation branch is applied for each modality respectively. The network structure used with the Squeeze and Excitation Fusion method is shown in Table 3.

Table 3: The network structure of the Squeeze and Excitation Fusion method. Conv means the convolutional layer, and each convolutional layer is followed by a batch normalization layer and a Rectified Linear Unit (ReLU). GAP stands for global average pooling. All the convolutional filters of three convolutional blocks are 3*3.

Layer	Kernel Size/Stride	Output Channels	Input
Conv1_1	7*7/4	32	**NIR img**
Conv2_1	3*3/1	64	Conv1_1
Blocks1_1	-	128	Conv2_1
Blocks2_1	-	256	Blocks1_1
Conv1_2	7*7/4	32	**VIS img**
Conv2_2	3*3/1	64	Conv1_2
Blocks1_2	-	128	Conv2_2
Blocks2_2	-	256	Blocks1_2
Features1 = concatenate(Blocks2_1, Blocks2_2)			
SE	-	512	Features1
Features2 = Features1 * SE			
Blocks3	-	64	Features2
GAP	-	64	Blocks3
Fc1	-	2	GAP2

The softmax loss function is adopted as the loss function. In the training phase, all models are trained in 5 epochs and optimized by the Stochastic Gradient Descent (SGD) algorithm. We set the momentum factor of the optimizer as 0.9. The learning rate was initially set to 0.1, and then it decays by 0.1 in every two epochs. We train all the models with the batch size 128 on 1 TITAN X (Pascal) GPU. Specifically, the classification confidence score of the video is the mean of the classification confidence score for each frame in the video.

5. Experiments

5.1. Evaluation Metrics

In the following experiments, we adopt four indices, i.e., equal error rate (EER), attack presentation classification error rate (APCER), bona fide presentation classification (BPCER) and average classification error rate (ACER) [10]. The threshold corresponding to the APCER and BPCER computation in the test set is set as the one corresponding to the EER in the validation set.

5.2. Results and Analysis

The experimental results evaluated with Protocol 1 are shown in Table 4. This result shows that the Squeeze and Excitation Fusion method performs the best (ACER=2% and EER@Test=1.8%) when the scene of the training set and the test set are consistent. The performance of algorithms using only single-mode data is similar, and the CNN-based algorithms perform better. Because the CNN-based algorithms generally have a better learning ability, and the Squeeze and Excitation Fusion method could fully use bi-modal data.

Then we explore the performance of the baseline algorithms under the influence of illumination changes. From Table 4 and Table 5, we can see that the ACER and EER of most baseline methods is getting worse when illumination changes. From Table 5, we can see that the Naive Data Fusion method (ACER=5.5%, 2% and EER@Test=1.4%, 1.3%) and the LBP-based algorithm using only near-infrared modal data (ACER=5.5%, 2.7% and EER@Test=5.5%, 4.1%) achieve the best performance. The algorithms using only NIR samples are better than the ones only using the VIS samples. It indicates that the visible lighting imaging is more sensitive to the illumination variation, while the NIR image is more robust to environmental illumination changes. Moreover, the deep learning methods are subject to over-fitting and require a high amount of data, which perform relatively poor. It is worth noting that the Naive Data Fusion method increases the amount of data in disguise, so the over-fitting problem is alleviated and its evaluation result is relatively better.

Finally, we focus on the impact of distance. From Table 6-9, we can conclude: 1) The CNN-based algorithm using VIS images has the best generalization across different distances; 2) When using only NIR images, the LBP-based method is much better than the CNN-based method; 3) The Naive Data Fusion method also achieves good results, such as ACER=9.2% (Table 8), 8.3% (Table 9). Moreover, comparing to the NIR images, the VIS image is more robust to different distances. The generalization across different distances of all methods is not good enough for real world applications, more powerful anti-spoofing methods have to be designed in future.

In summary, several conclusions can be drawn from the above experiments following three protocols. From the above experiments we can see that both illumination and distances can significantly influent the anti-spoofing performance. Due to the limitation of training data, CNN-based methods not always perform better than LBP-based methods. In general, traditional hand-crafted features have robust performance when the training data is limited. The bi-modal fusion methods have outstanding effects in many aspects indicating that the VIS image and the NIR images can provide complementary texture and imagery information to highlight the difference between the genuine and the fake samples.

6. Conclusion

In this paper, we release a novel 3D mask anti-spoofing database containing multi-spectrum data, including the samples captured in the visible band and the near-infrared band. Specifically, this collected database contains 920 video samples consisting of 67 genuine subjects and 48 fabricated 3D masks isolated to the genuine persons. The video samples in the collected database are recorded in two illumination conditions and four distance conditions, which have a great impact on the anti-spoofing performance. We also propose three test protocols to evaluate the generalization capabilities of the developed face anti-spoofing algorithms across different acquisition conditions. Benchmark experiments using the convolutional neural network and the LBP-based face anti-spoofing algorithm have been performed on the proposed database to inspect the generalization capability of the methods. The lack of publicly available data is still an important reason hindering the development of face presentation attacks detection technology. We would like to invite the biometrics research community to collect and share more face anti-spoofing database.

Acknowledgements

This work was supported by the Chinese National Natural Science Foundation Projects #61876178, #61806196, #61872367, #61572501.

References

[1] Z. Boulkenafet, J. Komulainen, L. Li, X. Feng, and A. Hadid. OULU-NPU: A mobile face presentation attack database with real-world variations. In *2017 IEEE International Conference on Automatic Face and Gesture Recognition*, pages 612–618, 2017.

[2] M. M. Chakka, A. Anjos, and S. Marcel. Motion-based counter-measures to photo attacks in face recognition. *IET Biometrics*, 3(3):147–158, sep 2014.

[3] I. Chingovska, N. Erdogmus, A. Anjos, and S. Marcel. Face recognition systems under spoofing attacks. In *Face*

Table 4: The experimental results (%) evaluated with Protocol I.

Modality	Method	EER@Val	BPCER	APCER	ACER	EER@Test
NIR	ResNet-9	2.0	6.2	**0.6**	3.4	4.1
	LBP-SVM	1.1	4.8	2.0	3.4	4.5
VIS	ResNet-9	3.7	**2.7**	6.9	4.8	4.6
	LBP-SVM	8.9	4.8	7.6	6.2	6.6
NIR + VIS	Concatenation Data Fusion	1.4	4.8	6.9	5.9	6.5
	Naive Data Fusion	**0.0**	4.8	2.0	3.4	3.8
	Squeeze and Excitation Fusion	0.7	**2.7**	1.3	**2.0**	**1.8**

Table 5: The performance (%) of the algorithms under the influence of illumination changes.

Modality	Method	Light: 200 lux					Light: 500 lux				
		EER@Val	BPCER	APCER	ACER	EER@ Test	EER@Val	BPCER	APCER	ACER	EER@Test
NIR	ResNet-9	5.8	15.2	2.7	9	11.1	2.2	12.5	0	6.2	1.8
	LBP-SVM	**2.2**	6.9	1.3	**4.1**	**5.5**	2.2	4.1	1.3	2.7	4.1
VIS	ResNet-9	10.4	0	34.7	17.3	12.5	3.9	12.5	31.9	22.2	25.9
	LBP-SVM	12.6	8.3	20.8	14.5	15.5	2.9	13.8	2.7	8.3	3.8
NIR + VIS	Concatenation Data Fusion	6.9	0	27.7	13.8	9.7	4.4	20.8	43	31.9	30.5
	Naive Data Fusion	5.8	5.5	6.9	6.2	**5.5**	**1.4**	1.3	2.7	**2**	**1.3**
	Squeeze and Excitation Fusion	3.6	0	20.8	10.4	6.2	**1.4**	19.4	13.8	16.6	15.9

Table 6: Generalization (%) across different distances. The distance setting of the train set is 30cm.

Modality	Method	EER@Val	BPCER	APCER	ACER	EER@Test
NIR	ResNet-9	5.6	17.5	58.3	37.9	33.3
	LBP-SVM	**0**	67.5	0	33.7	**2.7**
VIS	ResNet-9	8.8	3.7	13.8	**8.7**	6
	LBP-SVM	15.5	12.9	6.4	9.7	6.9
NIR + VIS	Concatenation Data Fusion	5	2.7	72.2	37.5	10.4
	Naive Data Fusion	5.4	12	19.4	15.7	16.2
	Squeeze and Excitation Fusion	**0**	2.7	62	32.4	12

Table 7: Generalization (%) across different distances. The distance setting of the train set is 60cm.

Modality	Method	EER@Val	BPCER	APCER	ACER	EER@Test
NIR	ResNet-9	**0**	48.1	4.6	26.3	18.5
	LBP-SVM	**0**	4.6	9.2	**6.9**	**8.7**
VIS	ResNet-9	**0**	12	17.5	14.8	14.8
	LBP-SVM	7.8	13.8	2.7	8.3	5.1
NIR + VIS	Concatenation Data Fusion	**0**	18.5	19.4	18.9	19.1
	Naive Data Fusion	**0**	17.5	13.8	15.7	14.7
	Squeeze and Excitation Fusion	**0**	16.6	18.5	17.5	17.1

Recognition Across the Imaging Spectrum, pages 165–194. Springer, 2016.

[4] 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 *2013 International Conference on Biometrics*, pages 1–8. IEEE, 2013.

[5] N. Erdogmus and S. Marcel. Spoofing in 2D face recognition with 3D masks and anti-spoofing with Kinect. *2013 IEEE International Conference on Biometrics: Theory, Applications and Systems*, pages 1–6, 2013.

[6] J. Guo, X. Zhu, Z. Lei, and S. Z. Li. Face synthesis for eyeglass-robust face recognition. In *Chinese Conference on*

Table 8: Generalization (%) across different distances. The distance setting of the train set is 90cm.

Modality	Method	EER@Val	BPCER	APCER	ACER	EER@Test
NIR	ResNet-9	2.9	62.9	0.9	31.9	33.7
	LBP-SVM	**0.0**	7.4	9.2	8.3	8.3
VIS	ResNet-9	1.4	8.3	5.5	**6.9**	**6.4**
	LBP-SVM	13.2	10.1	16.6	13.4	12.9
NIR + VIS	Concatenation Data Fusion	**0.0**	55.5	4.6	30.0	34.2
	Naive Data Fusion	**0.0**	15.7	2.7	9.2	10.6
	Squeeze and Excitation Fusion	**0.0**	50.9	9.2	30.0	34.0

Table 9: Generalization (%) across different distances. The distance setting of the train set is 120cm.

Modality	Method	EER@Val	BPCER	APCER	ACER	EER@Test
NIR	ResNet-9	5.2	87.9	0.0	43.9	39.0
	LBP-SVM	**0.0**	11.1	0.9	6.0	5.0
VIS	ResNet-9	4.4	6.4	1.8	**4.1**	**3.0**
	LBP-SVM	9.4	11.1	11.1	11.1	11.1
NIR + VIS	Concatenation Data Fusion	2.9	81.4	0.0	40.7	35.1
	Naive Data Fusion	2.9	14.8	1.8	8.3	7.4
	Squeeze and Excitation Fusion	3.9	86.1	0.0	43.0	34.5

Biometric Recognition, pages 275–284. Springer, 2018.

[7] J. Guo, X. Zhu, J. Xiao, Z. Lei, G. Wan, and S. Z. Li. Improving face anti-spoofing by 3d virtual synthesis. *arXiv preprint arXiv:1901.00488*, 2019.

[8] J. Hu, L. Shen, and G. Sun. Squeeze-and-Excitation Networks. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7132–7141. IEEE, jun 2018.

[9] G. B. Huang, M. Mattar, T. Berg, and E.Learned-Miller. Labeled faces in the wild: A database forstudying face recognition in unconstrained environments. In *Workshop on faces in'Real-Life'Images: detection, alignment, and recognition*, 2008.

[10] ISO/IEC JTC 1/SC 37 Biometrics. Information technology - Biometric presentation attack detection Part 1: Framework, 2016.

[11] A. Jourabloo, Y. Liu, and X. Liu. Face de-spoofing: Anti-spoofing via noise modeling. *ECCV 2018*, 11217 LNCS:297–315, 2018.

[12] I. Kemelmacher-Shlizerman, S. M. Seitz, D. Miller, and E. Brossard. The megaface benchmark: 1 million faces for recognition at scale. In *ICCV*, 2016.

[13] J. Komulainen, A. Hadid, and M. Pietikainen. Context based face anti-spoofing. In *2013 IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems*, pages 1–8. IEEE, sep 2013.

[14] J. Liu and A. Kumar. Detecting presentation attacks from 3d face masks under multispectral imaging. In *2018 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 47–475, June 2018.

[15] S. Liu, B. Yang, P. C. Yuen, and G. Zhao. A 3D mask face anti-spoofing database with real world variations. In *2016 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pages 1551–1557, 2016.

[16] Y. Liu, A. Jourabloo, and X. Liu. Learning deep models for face anti-spoofing: binary or auxiliary supervision. In *2018 IEEE Conference on Computer Vision and Pattern Recognition*, pages 389–398. IEEE, 2018.

[17] K. Patel, H. Han, and A. K. Jain. Secure Face Unlock: Spoof Detection on Smartphones. *IEEE Transactions on Information Forensics and Security*, 11(10):2268–2283, oct 2016.

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

[19] L. Wolf, T. Hassner, and I. Maoz. Face recognition in unconstrained videos with matched background similarity. In *CVPR 2011*, pages 529–534, 2011.

[20] D. Yi, Z. Lei, Z. Zhang, and S. Z. Li. Face anti-spoofing: Multi-spectral approach. In *Advances in Computer Vision and Pattern Recognition*, volume 49, pages 83–102. Springer, 2014.

[21] S. Zhang, X. Wang, A. Liu, C. Zhao, J. Wan, S. Escalera, H. Shi, Z. Wang, and S. Z. Li. A dataset and benchmark for large-scale multi-modal face anti-spoofing. In *CVPR 2019*, 2018.

[22] S. Zhang, X. Zhu, Z. Lei, H. Shi, X. Wang, and S. Z. Li. Faceboxes: A CPU real-time face detector with high accuracy. In *2017 IEEE International Joint Conference on Biometrics*, pages 1–9. IEEE, oct 2017.

[23] Z. Zhang, J. Yan, S. Liu, Z. Lei, D. Yi, and S. Z. Li. A face antispoofing database with diverse attacks. In *2012 IAPR International Conference on Biometrics*, pages 26–31, 2012.

[24] Z. Zhang, D. Yi, Z. Lei, and S. Z. Li. Face liveness detection by learning multispectral reflectance distributions. In *2011 IEEE International Conference on Automatic Face and Gesture Recognition and Workshops*, pages 436–441. IEEE, 2011.