

DOCTORAL THESIS

3D Mask Face Presentation Attack Detection with Remote Photoplethysmography

LIU, Siqi

Date of Award:
2021

[Link to publication](#)

General rights

Copyright and intellectual property rights for the publications made accessible in HKBU Scholars are retained by the authors and/or other copyright owners. In addition to the restrictions prescribed by the Copyright Ordinance of Hong Kong, all users and readers must also observe the following terms of use:

- Users may download and print one copy of any publication from HKBU Scholars for the purpose of private study or research
- Users cannot further distribute the material or use it for any profit-making activity or commercial gain
- To share publications in HKBU Scholars with others, users are welcome to freely distribute the permanent URL assigned to the publication

Abstract

Since face recognition has been widely employed in a variety of applications, including e-commerce and access control of unmanned security doors and mobile phones, security issues of face recognition systems have received increasing attention. Face presentation attack (PAD) is one of the greatest challenges in practical face recognition systems since images or videos of a user’s face can be easily acquired and printed. With the advancement of 3D printing and material technologies, super-real facial masks can successfully spoof existing face recognition systems at an affordable cost. Several approaches have been proposed to exploit appearance differences between masks and real faces. Although encouraging results are reported on single dataset, they can hardly generalize across different types of masks and record settings in real application scenarios.

In this thesis, we propose to tackle 3D mask face PAD by analyzing facial heartbeat signals through the remote photoplethysmography (rPPG) technique. Such a liveness cue is intrinsic to characterizing a true face and therefore can be used to detect a 3D mask, regardless of the material and quality of the mask. We first build a new 3D mask attack dataset called HKBU-MARs to simulate the real-world variations including different mask types, lighting conditions, and camera settings. For using rPPG for 3D mask face PAD, we first propose a local solution which extracts an rPPG correlation pattern from multiple local facial regions and learn the corresponding confidence map to encode the robust spatial information of heartbeat strength. To identify the heartbeat vestige from the observed noisy rPPG signals, rPPG correspondence feature with the noise-aware template learning and verification framework is developed. The two rPPG-spectrum-based methods require long time observation to identify the heartbeat information. To shorten the observation time, we further propose a fast rPPG-based solution by analyzing the similarity of local facial rPPG signals in the time domain. Furthermore, a spatiotemporal-convolution-based rPPG estimator for general use is designed, which could be a future direction of boosting the performance of rPPG-based 3D mask face PAD.

Keywords: Face Presentation Attack Detection, 3D Mask Attack, Remote Photo-plethysmography

Table of Contents

Declaration	i
Abstract	ii
Acknowledgements	iv
Table of Contents	vi
List of Tables	x
List of Figures	xii
Chapter 1 Introduction	1
1.1 Background	1
1.1.1 2D Face Presentation Attack Detection	1
1.1.2 3D Mask Face Presentation Attack Detection	2
1.2 Thesis Goals and Contributions	3
1.3 Related Work	4
1.3.1 Appearance-Based Face Presentation Attack Detection	4
1.3.2 Motion-Based Face Presentation Attack Detection	5
1.3.3 Other Liveness Cues	5
1.4 Organization	6
Chapter 2 HKBU-MARs: A New 3D Mask Face Presentation Attack Detection Database with Real World Variations	7
2.1 Introduction	7
2.2 Dataset Construction	8
2.2.1 Mask types	9
2.2.2 Cameras	10
2.2.3 Lighting conditions	11
2.2.4 Recording settings	12

2.3	Testing Protocols	13
2.3.1	Intra-Variation Test	13
2.3.2	Cross-Variation Test	13
2.4	Experiments	14
2.4.1	Baseline methods	14
2.4.2	Results and Analysis	14
2.5	Discussion	16
2.6	Summary	17

Chapter 3 3D Mask Face Presentation Attack Detection with Local Remote Photoplethsmography 19

3.1	Introduction	19
3.2	Principle of rPPG Based 3D Mask Presentation Attack Detection (PAD)	21
3.2.1	Why rPPG Works for 3D Mask Face PAD	21
3.2.2	Local rPPG for 3D Mask PAD	22
3.3	Proposed Method	23
3.3.1	Local rPPG Extraction	24
3.3.2	Local rPPG Correlation Model	24
3.3.3	Learning Local rPPG Confidence Map	27
3.3.4	Classification	29
3.3.5	Implementation Details	30
3.4	Experiments	30
3.4.1	Experimental Settings	30
3.4.2	Intra-Dataset Evaluation	33
3.4.3	Cross-Dataset Evaluation	35
3.4.4	Cross Mask Evaluation	36
3.4.5	Insensitivity to Training Subject	37
3.4.6	Convergence of Local Confidence Map Learning	39
3.5	Summary	40

Chapter 4 Find the True Liveness Sign from Noisy Signal: rPPG Correspondence Feature for 3D Mask Face PAD 41

4.1	Introduction	41
4.1.1	Limitation of Previous Local rPPG Solution	41
4.1.2	Contribution of This Chapter	42
4.2	Analysis of rPPG-based Face PAD	43
4.3	Multi-Channel rPPG Correspondence Feature for 3D Mask Face PAD	45

4.3.1	Preprocessing	45
4.3.2	rPPG Correspondence Feature for 3D Mask Face PAD (CFrPPG)	46
4.3.3	Multi-Channel CFrPPG with Time-Frequency Analysis	49
4.3.4	Implementation Details	52
4.4	Experiments	52
4.4.1	Datasets	52
4.4.2	Experimental Settings	55
4.4.3	Intra-Dataset Evaluation	56
4.4.4	Cross-Dataset Evaluation	59
4.4.5	Robustness to Different Masks Types and Transmittance	62
4.4.6	Robustness to Global Noise in More Practical Scenarios	62
4.5	Summary	63
Chapter 5 Fast 3D Mask Face Presentation Attack Detection with rPPG		66
5.1	Introduction	66
5.1.1	Limitation of rPPG-Spectrum-based 3D mask PAD	66
5.1.2	Contribution of This Chapter	67
5.2	Analysis of rPPG-Spectrum-based 3D mask PAD	67
5.3	Proposed Method	69
5.3.1	Temporal rPPG Similarity Feature for Fast 3D Mask PAD	69
5.3.2	Classification	73
5.4	Experiments	73
5.4.1	Experimental Setup	73
5.4.2	Intra-Dataset Evaluation	74
5.4.3	Ablation Study	75
5.4.4	Cross-Dataset Evaluation	76
5.4.5	Robustness to Lighting Variation	79
5.4.6	Robustness to Different Masks Transmittance and Eyeglasses Occlusion	82
5.5	Summary	83
Chapter 6 A General rPPG Estimator with Spatiotemporal Convolutional Network		84
6.1	Introduction	84
6.1.1	Limitation of Existing rPPG Estimator	84
6.1.2	Contribution of This Chapter	85
6.2	Related Work	86

6.2.1	Remote Photoplethysmography Estimation	86
6.2.2	Spatiotemporal Convolution	87
6.3	Proposed Method	87
6.3.1	Spatiotemporal Convolution for rPPG measurement	87
6.3.2	Noise-Aware Robust Triplet-DeepPPG	91
6.3.3	Spatiotemporal rPPG aggregation	92
6.4	Experiments	94
6.4.1	Experimental Settings	94
6.4.2	Cross-skin Evaluation of Heart Rate Estimation	96
6.4.3	Evaluation of Heart Rate Estimation	97
6.4.4	DeepPPG for rPPG-based 3D mask Presentation Attack De- tection	100
6.5	Summary	101
Chapter 7 Conclusion and Future Work		104
7.1	Conclusion	104
7.2	Future Research Directions	105
7.2.1	Real-time rPPG-based 3D Mask Presentation Attack Detec- tion via Deep Learning	106
7.2.2	Motion Robust rPPG-based 3D Mask Face PAD	106
7.2.3	Compression Robust rPPG-based 3D Mask Face PAD	107
Bibliography		108
List of Publications		115
CURRICULUM VITAE		116

List of Tables

2.1	Variation summary of 3D mask attack datasets used in the experiment	8
2.2	The EER(%) of intra-variation experiment on two types of mask attack	15
2.3	The EER(%) of cross lighting experiment	15
2.4	The EER(%) of cross camera experiment	15
2.5	The EER(%) of cross mask type experiment	16
3.1	Comparison results under intra dataset protocol on the 3DMAD dataset	34
3.2	Comparison results under intra dataset protocol on the HKBU-MARsV1+ dataset	34
3.3	Comparison results under intra dataset protocol on the Combined dataset	34
3.4	Cross-dataset evaluation results between 3DMAD and HKBU-MARsV1+ datasets.	38
3.5	Cross-mask evaluation results on the HKBU-MARsV1+ dataset. . . .	38
3.6	Cross-mask evaluation results on the Combined dataset.	38
4.1	Variation summary of 3D mask attack datasets used in the experiment	53
4.2	Intra dataset evaluation results(%) on HKBU-MARsV2+	56
4.3	Intra dataset evaluation results(%) on HKBU-MARsV1+	57
4.4	Intra dataset evaluation results(%) on 3DMAD	57
4.5	Comparison results(%) under intra dataset protocol on CSMAD . . .	60
4.6	LOLO and LOCO evaluation results(%) on HKBU-MARsV2+	60
4.7	Cross-dataset evaluation results between 3DMAD, HKBU-MARsV1+, HKBU-MARsV2+, and CSMAD. $A \Leftrightarrow B$ indicates the evaluation across datasets A and B, where the left column is $A \rightarrow B$ and right one is $B \rightarrow A$. HTER standard deviation is in bracket.	61
4.8	Ablation study results(%) on Replay Attack Dataset	63
5.1	Intra dataset evaluation results on 3DMAD with short observation time (1 second)	75

5.2	Intra dataset evaluation results on HKBU-MARsV1+ with short observation time (1 second)	75
5.3	Performances (AUC) with different length of observation time.	76
5.4	LOOCV and LOVO evaluation results on HKBU-MARsV2+-indust with short observation time (1 second)	76
5.5	Cross-dataset evaluation results between 3DMAD and HKBU-MARsV1+ with short observation time (1 second)	77
5.6	Cross-dataset evaluation results between 3DMAD and HKBU-MARsV2+ with short observation time (1 second)	78
5.7	Cross-dataset evaluation results between HKBUMARsV1+ and HKBU-MARsV2+ with short observation time (1 second)	78
5.8	Cross-dataset evaluation results between CSMAD and 3DMAD with short observation time (1 second)	80
5.9	Cross-dataset evaluation results between CSMAD and HKBU-MARsV1+ with short observation time (1 second)	80
5.10	Cross-dataset evaluation results between CSMAD and HKBU-MARsV2+ with short observation time (1 second)	80
5.11	Evaluation on CSMAD with short observation time (1 second)	83
6.1	Cross-skin evaluation of DeeprPPG on PURE dataset. A→B indicates train on skin region A and test on skin region B. F, C, and L refer to forehead, cheek, and lower face, respectively.	96
6.2	Evaluation of average (HR) measurement on PURE dataset	98
6.3	Evaluation of average heart rate on COHFACE dataset	98
6.4	Evaluation of average heart rate on MAHNOB-HCI dataset	99
6.5	Intra-dataset evaluation for 3D mask face PAD on 3DMAD	101
6.6	Intra-dataset evaluation for 3D mask face PAD on HKBU-MARsV1+	101
6.7	Cross-dataset evaluation for 3D mask face PAD between 3DMAD and HKBU-MARsV1+	102

List of Figures

2.1	Sample mask images in the proposed new 3D mask face PAD database. (a)-(f) are ThatsMyFace masks and (g)-(l) are REAL-f masks.	9
2.2	High resolution sample images of Thatsmyface (TF) mask (a) and REAL-f mask (b).	10
2.3	Sample face images recorded by different cameras under same lighting condition (room light).	11
2.4	Sample face images recorded under different lighting conditions. Images are captured by Sony Tablet S.	12
2.5	ROC curves of three baseline methods under overall testing protocol .	16
3.1	Effect of remote photoplethysmography (rPPG) on (a) a normal unmasked face, and (b) a masked face. (a) shows rPPG on a genuine face: sufficient light penetrates the semi-transparent skin tissue and interact with blood vessels. The rPPG signal can go through skin and be detected by an RGB camera. (b) depicts rPPG on a masked face: the mask material blocks a large portion of the light that the skin should absorb. The light source needs to penetrate a layer of painted plastic and a layer of skin before interacting with the blood. Any remaining rPPG signals will be too weak to be detected	20
3.2	The distribution of local rPPG signal strength on genuine faces for different subjects. The value of each pixel is obtained through densely extract local rPPG signals and calculate their signal to noise ratios. The color from red to blue demonstrates the signal strength from strong to weak.	23

3.3	Block diagram of the proposed LrPPG method. It consists four main components: (1) local rPPG extraction; (2) local rPPG correlation modeling; (3) learning local rPPG confidence map and (4) classification. For an input face video, local rPPG signals are extracted from the local facial regions defined by landmarks. Then the local rPPG correlation pattern is extracted from the input signals as the liveness feature. In training stage, the local rPPG confidence map of local facial region is learned and transformed into distance metric in classification. Finally, classifier is trained to detect the input testing local rPPG correlation pattern.	25
3.4	Effect of local rPPG cross correlation on genuine face and masked face. The first two rows are the PSD of two local rPPG signals \mathbf{s}_i and \mathbf{s}_j and the last row is the cross correlation result of them. Here (a) is the cross-correlation of two strong local rPPG signals, (b) is of one strong and one noisy signal, and (c) is the cross-correlation of two signals on a masked face.	26
3.5	The LOF caption	31
3.6	Average ROC curves of three datasets under intra-dataset protocol. .	34
3.7	AUC under intra-dataset protocol with different number of training subjects.	35
3.8	Average ROC curves under cross-dataset protocol.	35
3.9	AUC under cross-dataset protocol with different number of training subjects.	36
3.10	Average ROC curves under cross-mask protocol on the HKBU-MARsV1+ dataset. TMF and RF represent Thatsmyface and REAL-f, respectively.	37
3.11	AUC under cross-mask protocol on the MARsV1+ dataset with different numbers of training subjects. TMF and RF represent Thatsmyface and REAL-f, respectively.	37
3.12	Average ROC curves under cross-mask protocol on the Combined dataset.	39
3.13	AUC under cross-mask protocol on the Combined dataset with different numbers of training subjects.	39
3.14	Convergence of confidence map learning under cross-dataset protocol	40

4.1	Three typical rPPG signal patterns in 3D mask presentation attack detection (PAD). Ideally, the difference of rPPG signals from genuine face and masked face is significant. However, the rPPG signal is fragile to interference in practical scenarios	44
4.2	Block diagram of the proposed Multi-Channel rPPG Correspondence Feature (MCCFrPPG)	46
4.3	Example face images from the 4 3D mask face anti-spoofing datasets: (a) 3DMAD, (b) HKBU-MARsV1+, and the extended dataset (c) HKBU-MARsV2+. The first row shows the genuine faces in 6 lighting conditions and the second row shows the hyper real mask examples made by REAL-f (left four are newly added). Images are captured through the industrial camera with fix exposure rate and focus. (d) CSMAD that contains glasses occlusion and severe side light.	54
4.4	Average ROC curves of CSMAD under intra-dataset testing protocol. There are 4 lighting conditions (bright light, room light, side light) in CSMAD dataset. Result in (a) includes all lighting conditions for training and testing while (b) is trained on bright light and tested on the other three.	58
4.5	Average ROC curves of HKBU-MARsV2+ under LOOCV, leave one lighting condition out (LOLO). and leave one camera out (LOCO) protocol.	58
4.6	Average ROC curves of RAD datasets under intra-dataset protocol. CFrPPG ^{-NAR} indicates the single channel CFrPPG without noise-aware component.	63
4.7	Average ROC curves of all cross dataset combinations between 3DMAD, HKBU-MARsV1+, HKBU-MARsV2+, and CSMAD datasets. Note that A→B indicates the setting of training on dataset A and testing on dataset B.	65
5.1	An example of rPPG frequency analysis when the observation time becomes short. The right subfigure visualizes the spectrum variation when the observation time becomes shorter. Each row in the matrix represents the average of multiple local rPPG spectrums given a specific observation time length.	68
5.2	Similarity comparison of local rPPG signals between genuine face and masked face, in terms of the amplitude, gradient and phase.	69
5.3	Example of TSrPPG matrix on genuine face (left two) and masked face (right two)	72

5.4	Average ROC curves under cross-dataset testing between 3DMAD and HKBUMARsV2+ with short observation time (1 second)	79
5.5	Average ROC curves under cross-dataset testing between HKBUMARsV1+ and HKBUMARsV2+ with short observation time (1 second)	81
5.6	Average ROC curves under cross-dataset testing between CSMAD and 3DMAD with short observation time (1 second)	81
5.7	Average ROC curves under cross-dataset testing between CSMAD and HKBUMARsV1+ with short observation time (1 second)	81
5.8	Average ROC curves under cross-dataset testing between CSMAD and HKBUMARsV2+ with short observation time (1 second)	82
5.9	Average ROC in log scale on HKBU-MARsV2+ using LOOCV and LOVO protocol with 1-second observation	82
6.1	The proposed DeeprPPG is an rPPG estimator for general use. It can be applied to different input skin regions in wide application scenarios. The DeeprPPG is performed spatiotemporally on the video clip of the selected region to estimate the rPPG signal.	85
6.2	The proposed DeeprPPG architecture. DeeprPPG net includes 1 2D convolution, 7 spatiotemporal convolutions and 4 spatial average pooling layers followed by the final channel aggregation. The first 2D convolution kernels are 5×5 with stride 1×1 and all spatiotemporal convolutions kernels are $3 \times 3 \times 3$ with stride $1 \times 1 \times 1$. The number of channels (filters) is denoted in each box. SpatioPool1 to SpatioPool3 are performed spatially with kernel $1 \times 2 \times 2$, stride $1 \times 2 \times 2$. SpatioPool4 is an adaptive average pooling that squeezes spatial feature area into 1. The final rPPG is obtained through the channel aggregation.	88
6.3	The demonstration of STConv blocks. The details of spatiotemporal blocks in Figure 6.2 is shown in (a) which includes the omitted 3D batch norm and ReLU. (b) Full 3D convolution with a kernel size $3 \times 3 \times 3$, where the first dimension is time and rests are width and height. (c) The (2+1)D block separates the full 3D convolution into a spatial 3D convolution with a temporal 1D convolution.	90
6.4	Noise-aware robust Triplet-DeeprPPG network	91

6.5 The illustration of spatiotemporal rPPG aggregation. Given an input video with M multiple ROIs, the region is warped into $W \times H$ over frames and form the ROI videos. DeepPrPPG is consecutively performed on ROI clip $T \times W \times H$. The resulting rPPG fragments are weighted summarized into the final rPPG signal according to their reliability. 93

6.6 Example of different face status in PURE datasets: (a) stable, (b) talking, (c) slow head translation, and (d) medium head rotation. Note that the talking and head rotation can cause severe color, light, or proportion variation on facial skin, which are challenging cases in rPPG estimation. The fast head translation and small head rotation are not shown here. 95

6.7 Example of two lighting conditions in COHFACE datasets: (a) bright studio light under 400W halogen spotlight and ceiling light, (b) side light by operating the natural light with blinds. 95

6.8 Visualization of our estimated rPPG signals under different face status in the PURE dataset: (a) stable, (b) talking, (c) slow head translation, and (d) medium head rotation. Each status contains samples from 4 different subjects 103