Detection of 3D face masks with thermal infrared imaging and deep learning techniques

Marcin Kowalski,*[1] Krzysztof Mierzejewski[2]

[1]*Institute of Optoelectronics, Military University of Technology, Gen. S. Kaliskiego 2, 00-908 Warsaw, Poland*
[2]*Faculty of Cybernetics, Military University of Technology, Gen. S. Kaliskiego 2, 00-908 Warsaw, Poland.*

Received March 10, 2021; accepted June 30, 2021; published June 30, 2021

Abstract—Biometric systems are becoming more and more efficient due to the increasing performance of algorithms. These systems are also vulnerable to various attacks. Presentation of falsified identity to a biometric sensor is one of the most urgent challenges for the recent biometric recognition systems. Exploration of specific properties of thermal infrared seems to be a comprehensive solution for detecting face presentation attacks. This letter presents the outcome of our study on detecting 3D face masks using thermal infrared imaging and deep learning techniques. We demonstrate the results of a two-step neural network-featured method for detecting presentation attacks.

Face recognition systems are being challenged by several diverse malicious attacks. One of the most popular nowadays exploits someone else's identity presented to the biometric sensor. Face presentation attacks are relatively easy to carry out, even with printed photographs. More advanced presentation attack detection (PAD) methods are being developed in response to new presentation attacks introduced. The PAD method put forward herein is designed to utilize specific spectral features of thermal infrared imaging.

The aim of this paper is to present the outcomes of research into detection of various face presentation attacks in thermal infrared. An analysis of presentation attacks using novel 3D-printed and custom flexible 3D-latex masks is provided. The distribution of thermal radiation emitted by the face has been studied, particularly emission variation between specific landmarks of the face and neck. The paper presents the design of a PAD method together with validation results.

A variety of presentation attack detection methods have been proposed, mainly operating in the visible light domain [1–4]. Another pertinent attack predictor might be the display of subject heat emission, which is a distinctive quality of thermal infrared [5–7]. Still, on the grounds of apparatus availability, the visible light domain is so far predominant in current research, while the thermal infrared spectrum has received barely moderate attention.

Presentation attack might be detected with a thermal infrared camera as a result of comparing thermal emissions from bona fide with impostor faces. The thermal camera is a proper choice for a face presentation attack detector due to its ability to quantify the thermal energy of the subject. It is expected that heat emission of subject mounting presentation attack may change over the presence of the presentation attack instrument (PAI). The presentation attack instrument acts as an optical filter which limits the amount of energy reaching the imager.

The detection of face presentation attacks in thermal infrared has been addressed in several works, though mostly exploring multi-channel mode combined with a visible light range. Sun *et al.* [8] proposed a liveness detection approach based on thermal infrared and visible spectra. The detection method uses a canonical correlation analysis (CCA) between a visible and thermal face. The results show that the method obtains a live detection rate of 85.1% and 90.8%, including and excluding glasses with a false acceptance rate of 0.1%. Other methods include a thermal face-convolutional neural network (Thermal Face-CNN) with External Knowledge [9]. The external knowledge is based on the calculated temperature of a subject's face. The face liveness detection relies on absolute temperatures registered by a thermal camera fused with visible range images. Precise measurement of temperature with a thermal camera is difficult in an unconstrained environment. The proposed method obtained the best accuracy of 0.7918 with the recall of 0.7434 and precision of 0.8298. George *et al.* [10] proposed a Multi-Channel Convolutional Neural Network (MC-CNN) tested with grayscale, depth, infrared, and thermal infrared images. The proposed method was reported to obtain an Average Classification Error Rate (ACER) of 2.59% and 0.84% for thermal imaging and a combination of Grayscale, Depth, Infrared, and Thermal, respectively. This method obtains high results when using several channels together and should not be considered in a single-channel configuration.

Current efforts in the field of PAD algorithms are focused on new methods that will accurately cover known and unknown attacks. Generalization is one of the most significant challenges for current PAD algorithms, as they are biased towards the training data.

* E-mail: marcin.kowalski@wat.edu.pl, krzysztof.mierzejewski@wat.edu.pl

The presentation attacks addressed in this study are performed using two types of 3D facial masks. The first type corresponds to facial masks made of hard resin manufactured during the 3D printing process. These masks cover only the facial region; they are inflexible and thus do not fully adhere to facial shapes. The masks are customized, printed being based on a 3D model generated according to 2D photographs of the subject's face.

The second type of mask is a full-face mask made of flexible foam-latex, rendering realistic men faces. The latex masks cover the entire head together with the neck. The inner surface of masks is adhesive, hence the masks may hold well without additional support when put on. The masks are designed with holes in eyes, nose, and mouth locations. Sample images of facial masks are presented in Fig. 1.

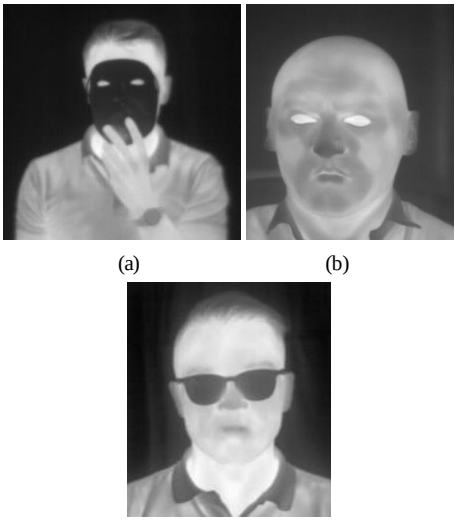
(a) (b)

Fig. 1. Images of the 3D-printed mask (a) and latex mask (b).

Thermal infrared imaging takes advantage of capturing the relative distribution of apparent temperature of objects located in the field of view. Quantitative analysis of thermal energy is defined by the noise equivalent temperature difference (NETD), which directly determines the camera ability to detect a slight difference in temperatures. During this study, a long-wavelength infrared camera operating in 7–14 μm has been used. The imager is equipped with an uncooled micro-bolometer focal plane array with NETD of 50 mK (at 300 K).

The range of PAIs used in this study includes facial as well as full-face masks. Since each PAI introduces a change of heat emission, it may be detected by differential analysis of a bare face and a covered face. Analysis of heat distribution should be then performed in the regions of interest containing the surface of the face and neighboring areas. Sample images presenting subjects wearing two types of masks are shown in Fig. 2. Analysis of collected thermal infrared images offering subjects wearing different masks in various configurations has led to the proposition of a method to detect 3D facial masks.

The proposed method detects attacks in a two-step manner. The first step of the algorithm is to detect the head and to determine the coordinates of a region of interest (RoI) corresponding to the head itself. In the second step, a trained classification algorithm performs the classification of the detected RoI. The head detection has been done using the Faster R-CNN [11] algorithm trained with thermal infrared and visible range face images. Faster R-CNN computes candidate regions by a fully convolutional region-proposal network. The Faster R-CNN algorithm may work with different parametrization networks. In this study, the ResNet-50 network [12] has been used for parametrization within the Faster R-CNN algorithm.



(a) (b)

(c)

Fig. 2. Sample images of presentation attacks; subject wearing 3D printed mask latex masks (a), a subject wearing the 3D-printed mask (b), genuine subject (c).

The region of interest is analyzed to detect differences between a real face and an attack. The head images extracted from original images are passed to a deep neural network for presentation attack detection. The analysis of facial heat maps to detect facial masks is proposed to be done with the ResNet-50 network. The ResNet network works as a typical classifier with two neurons at the final classification layer. Because the attack detector module was designed to operate fully autonomously, the optimizer loss function was set up to maximize the classification confidence score [13], subject to the clerical review area being empty in the training process. As both types of statistical errors are considered uniform in terms of cost, the decision boundary ensues, located precisely at 0 value of the log-likelihood ratio of attack to bona fide, which effectively renders the classifier follows the canonical Bayesian approach.

To achieve high detection performance, the classifier has been trained with a wide range of images of high variability. The CNN has been pre-trained on an ImageNet dataset containing visible domain images and fine-tuned on a variety of thermal images presenting actual samples, as well as spoofed samples. The subset

containing actual samples has been supplemented with images showing subjects after expending physical effort. Our analysis revealed that the subject's face could change significantly after a physical struggle. When the face is wet with sweat, the apparent temperature registered by a thermal imager decreases and starts to resemble the attacker's face.

The dataset has been divided into training, test, and validation sets with a split ratio of 70% (4900 images), 10% (700 images), and 20% (1400 images), respectively. Each part of the algorithm has been validated separately, beginning with the performance assessment of its first step, namely the RoI extraction component. The head detector is validated in terms of detection rate and false detection rate. Validation results of a head detector are presented in Table 1.

Table 1. Validation results of head detector[1].

Method	3D-printed		Latex	
	Det. Rate	False Det. Rate	Det. Rate	False Det. Rate
Faster R-CNN + ResNet 50	1.00	0.00	1.00	0.00

[1] Absolute values are given in the range between 0 and 1.

The following component has been validated with the next two PAD metrics: attack presentation classification error rate (APCER) and bona fide presentation classification error rate (BPCER). APCER corresponds to the proportion of attack presentations misusing the same PAI species classified as bona fide presentations in a specific scenario (False Negative Rate). At the same time, BPCER stands for the proportion of bona fide presentations incorrectly classified as attack presentations within a particular scenario (False Positive Rate).

For deep learning classifiers, we have applied two training and validation schemes. In the first one, a ten-fold cross-validation technique was applied. 70% of all images selected randomly were used as the training set (4900 images), while the remaining 30% constituted the testing and validation sets with a split ratio of 10% (700 images) and 20% (1400 images), respectively. The mean results of the 10-fold cross-validation are presented in Table 2.

Table 2. Results of 10-fold cross-validation.

Method	All	
	APCER[1]	BPCER[1]
ResNet-50	0.000	0. 000

[1] Absolute values are given in the range between 0 and 1.

For the second validation approach, the unknown-attack scenario was exercised to verify the proposed method generalization capacity. During the unknown-attack validation, training and testing splits were based on PAIs employed in the training phase. Under all the unknown-attack scenarios, images featuring attacks with one type of PAI were dispatched to the classifier for its training, and the remaining representing different PAIs were held out for testing. The latter group of images not applied to training is considered "unknown". The results of unknown-attack validation are presented in Table 3.

Table 3. Results of unknown-attack validation[1].

Method	3D-printed[2]		Latex[3]	
	APCER	BPCER	APCER	BPCER
ResNet-50	0.001	0.001	0.010	0.01

[1] Absolute values are given in the range between 0 and 1.

[2] Model trained on all masks except 3D-printed.

[3] Model trained on all masks except Latex masks.

The analysis of the results indicates that both steps of the algorithm achieve high performance. The head detection algorithm advertises zero error rate. The analysis of thermal distribution demonstrates impressive performance in the 10-fold cross-validation and unknown attack validation. The experiments in the unknown attack scenario show some disproportions between detection performance across different attacks, noting that a full-face latex mask disguise is more challenging to detect.

Thermal infrared imagers furnished with specific algorithms may be considered very efficient detectors of presentation attacks.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 787120.

References

[1] S.R. Arashloo, J. Kittler, W. Christmas, IEEE Trans. Inf. Forensics Secur. **10**, 11 (2015). https://ieeexplore.ieee.org/document/7163625

[2] A. Anjos, M.M. Chakka, S. Marcel, IET Biometrics **3,** 3 (2014). https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/iet-bmt.2012.0071

[3] M. Killioğlu, M. Taşkiran, N. Kahraman, Proc. SAMI IEEE, (2017). https://ieeexplore.ieee.org/document/7880281

[4] A. Asaduzzaman, A. Mummidi, M.F. Mridha, F.N. Sibai, Proc. ICAEE IEEE, (2015). https://ieeexplore.ieee.org/document/7506814

[5] M. Kowalski, Sensors **20**, 14 (2020). https://doi.org/10.3390/s20143988

[6] C. Galdi *et al*, IET Biometrics **9**, 6 (2020). https://doi.org/10.1049/iet-bmt.2020.0033

[7] D.A. Socolinsky, A. Selinger, J. Neuheisel, Comput. Vis Image Underst. **91**, 1 (2003). https://www.sciencedirect.com/science/article/pii/S1077314203000754?via%3Dihub

[8] L. Sun, W. Huang, M. Wu, *Proc.* CAIP, (2011). https://doi.org/10.1007/978-3-642-23678-5_12

[9] J. Seo, I. Chung, Symmetry 2019, **11**, 3 (2019). https://www.mdpi.com/2073-8994/11/3/360

[10] A. George, Z. Mostaani, D Geissenbuhler *et al*., IEEE Trans. Inf. Forensics Secur. **15**, (2020). https://ieeexplore.ieee.org/document/8714076

[11] S. Ren, K. He, R. Girshick, J. Sun, Proc. CVPR IEEE **39**, (2016). https://ieeexplore.ieee.org/document/517044

[12] K. He, X. Zhang, S. Ren, J. Sun, Proc. CVPR, (2016). https://ieeexplore.ieee.org/document/7780459

[13] K. Mierzejewski, M. Mazurek, Procedia Manufacturing **44**, 245-252 (2020). https://doi.org/10.1016/j.promfg.2020.02.228

