A. Brömme, C. Busch, A. Dantcheva, K. Raja, C. Rathgeb and A. Uhl (Eds.): BIOSIG 2020,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2020 1

Face Presentation Attack Detection in Ultraviolet Spectrum via Local and Global Features

Dirk Siegmund, Florian Kerckhoff, Javier Yeste Magdaleno, Nils Jansen,
Florian Kirchbuchner[1], Arjan Kuijper[2]

Abstract: The security of the commonly used face recognition algorithms is often doubted, as they appear vulnerable to so-called presentation attacks. While there are a number of detection methods that are using different light spectra to detect these attacks this is the first work to explore skin properties using the ultraviolet spectrum. Our multi-sensor approach consists of learning features that appear in the comparison of two images, one in the visible and one in the ultraviolet spectrum. We use brightness and keypoints as features for training, experimenting with different learning strategies. We present the results of our evaluation on our novel Face UV PAD database. The results of our method are evaluated in an leave-one-out comparison, where we achieved an APCER/BPCER of 0%/0.2%. The results obtained indicate that UV images in presentation attack detection include useful information that are not easy to overcome.

Keywords: Face Presentation Attack Detection PAD, Ultraviolet, MFP, Biometrics.

1 Introduction

Face recognition (FR) is the most commonly used biometric method for recognizing people. Applications range from unlocking smartphones and border-control to dynamic recognition in surveillance scenarios. The accuracy of face verification systems has improved significantly since the advent of deep learning, especially in scenarios where sample and probe image are taken in similar conditions. While the accuracy of FR improved, their vulnerability to presentation attacks remains a major challenge. Presentation attacks are defined as "presentation to the biometric data capture subsystem with the goal of interfering with the operation of the biometric system."[In16]. They range from very simple low effort attacks like printed face images or replayed videos to more sophisticated attacks involving high quality disguises and masks. Presentation attack detection (PAD) are approaches to prevent presentation attacks from single or series of images, using different properties like: motion, texture or life signs. There is currently no detection method that is absolutely safe. Especially three-dimensional masks and high quality 3D-prints can very often overcome PAD. Commonly known methods include additional analysis of images captured in different wavelengths, especially in the infrared (IR) and near-infrared (NIR). Their general vulnerability in practive is that 2D and 3D face images of people are commonly available, or can be generated even from a single image [ASJT17]. In this paper

[1] Fraunhofer Institute for Computer Graphics Research (IGD), Fraunhoferstrasse 5, 64283 Darmstadt, Germany {dirk.siegmund, florian.kerckhoff, javier.yeste.magdaleno, nils.jansen, florian.kirchbuchner}@igd.fraunhofer.de

[2] Technische Universität Darmstadt, Hochschulstr. 10, 64289 Darmstadt, Germany, arjan.kuijper@tu-darmstadt.de

we present a solution that tackles PAD, by using multi-modal biometrics in the ultra-violet (UV) spectrum by analyzing Melanin Face Pigmentation (MFP). As discovered in a recent paper [Sa18], MFP can be seen as additional modality, of which most can only been seen by a sensor, sensible in UV wavelength. In this paper we analyze if these captures of the human skin are useful for PAD by presenting novel methodology. Our first method detects MFP in the images using ORB keypoints and identifies attacks using their number and distribution. In the second method, we examine whether the corresponding brightness can be used as a feature between the images. Both methods use two captures at the same time, one in the UV spectrum and another made in the visual spectrum (VIS). To confirm our assumption that PAD works by using images in the UV spectrum, we present a database of presentation attacks that includes images in UV and VIS spectrum (see Section 3). This database contains images of 2D prints on paper, 3D prints and masks of different material. These images are evaluated together with a recently published database of 91 real subjects captured over a period of 6 months showing different expressions and poses. The methodology of our verification methods is presented in Section 4. There, we describe the image descriptors and fusion methodology that we used in our methodology. Our results in Section 5 show if UV face imaging and/or MFP, provide valuable distinct information for face PAD. We conclude with a future perspective about the use of these properties for future research and highlight observed issues and limitations in Section 6.

2 Related Work

Active imposter presentation attack detection algorithms can be categorized into hardware and software based. Software based algorithms are cheaper, space saving and include static and dynamic algorithms. They can analyze micro-textural patterns [RB17] and/or motion [De12] but mostly fail when a trained model is used in a different environment or on other datasets. Damer et al. [DD16] reported good results in a motion magnification based approach using histograms of oriented optical flow. A limitations of this approach is the human physiological rhythm itself and computational costs. Hardware-based multi-spectral algorithms analyze several images in distinct regions of the electro-magnetic spectrum individually [Ra17]. There are also multi-sensor approaches, where multiple spectral bands are being used at the same time by different sensors. The spectral band can be divided into the VIS [400nm - 700nm], IR [780nm - 15 μm], NIR and the short-wave (SWIR) band. Multi-sensor/cross model approaches can take advantage of the different reflection properties of material in different spectra. In other words, knowing that human skin reflects IR light quite different than e.g. silicon, enabled the detection presentation attacks by a comparison of images which capture both spectra. The effectiveness of this method is demonstrated by the known FaceID, used in the Apple iPhone's. But while active IR or NIR images show advantages especially in robustness to illumination and exhibit special characteristics of the human skin, they can be spoofed as well by using a 3D mask[SKJ16]. Due to the MFP ascribed properties, we think that these features should also be useful for PAD.

3 Database

The evaluation of the proposed method is carried out on an extended version of a newly created UV-Face database [Sa18]. The database consists of images collected in the UV, as well as in the VIS spectrum under conditions, as one would expect them in a controlled scenario, such as border control. Compared to the IR bandwidth, one of the first observations when exposing human skin to UV emission is, that skin of different people looks quite differently in that spectra. The Fitzpartick scale[Fi88] groups the skin type into six different categories, according to the reaction of the skin to the sun. Most notable with skin type I, where people show additional MFP in the UV image, that aren't visible for human eyes. We captured 476 images of 28 identities of Skintype I and II. 1042 images of 45 identities of skintype III and IV and 330 images of 18 identities of skintype V and VI. The database includes subjects of different age, gender and skin types. We've expanded the database by 127 images of spoofing attacks by using a variety of materials based on a selection of attacks according to reported attacks in media and research. We used eight different types of masks (painted and unpainted latex and latex foam), bursts (silicone, photopolymer and PLA) and paper printouts on different paper. Each attack is captured by using both cameras

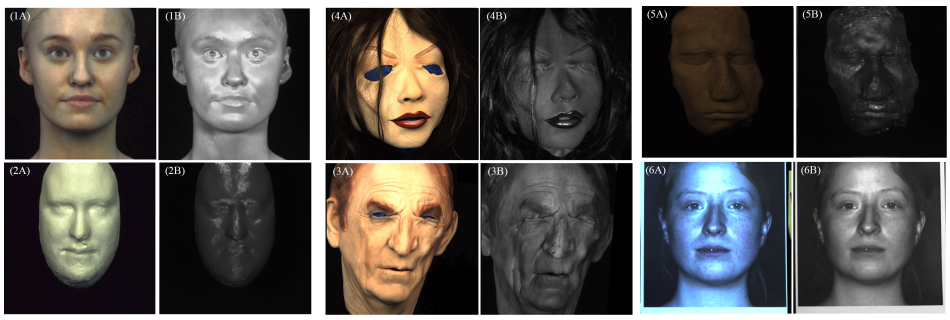

Fig. 1: Created spoofing attacks VIS (A) UV (B). (1) Color-bust made of photopolymers, by Stratasys - Connex 3 3D printer (Polyjet) (2) 3D face bust (17x11cm), by Prusa i3 MK3 3D (Polylactide) (3/4) Unpainted, professional latex masks, two painted masks of the same material and variations with wigs (5) 3D face bust (silicone rubber) on a 3D mold, using alginat for the imprint (6) Twenty laser color-printouts, Ten using normal paper, ten on thicker shiny

in following poses: frontal, 45°view to the left, 45°view to the right, looking up, looking down. Two cameras, attached side by side, are used in order to keep the divergence in perspective small. Test participants, wearing the masks, or the 3D models are positioned at a distance of 1.5m away from the cameras. In order to avoid interferences, UV/IR and VIS filters were used respectively, allowing only the transmission of the intended wavelength. For the UV capturing, a DLP LLC camera with a CMOS sensor, resulting in images of 2592x1944 pixel resolution is used. For illumination we used two 36W UV-A LPS lamps with a bandwidth between 315nm and 400nm positioned in front left and front right to the subject. The position of the used lights was chosen in a way that shades are similar in both captures. The images in the visible spectrum are captured by using a Nikon D9000 with a APS-C CMOS sensor and a 35mm lens. The UV images are resized by 58% and cropped to 600x600pixel, VIS images respectively. All images are converted to gray-scale.

We augmented the attack database by slightly changing the saturation for every image pair by using linear transformation.

4 Introduced Methods for UV-PAD

As one of the first observations, after capturing the attack images, we found that the brightness of the images differs greatly in UV compared to the VIS. Since both VIS and UV image are taken simultaneously by us, we can rule brightness manipulation by the attacker out. While the brightness of the silicone bust (see Figure 1-5B) is relatively low, the 3D color print made of photopolymers 1-1B) reflects a lot and is therefore very bright. Of course, it can also be assumed that UV images of non-skin have no MFP, which would be additionally evaluable on the UV images. Another observation is that relatively smooth material, such as latex masks with no notches, have almost no details in the UV spectrum (see Figure 1-4B). Furthermore, all latex masks show no reflections that lead to overexposure at all. Comparing that to bona fide images we observed that there is almost no image that does not show at least a small area like this (very often at the forehead). However, smooth material such as the silicon print, the 2D prints or the PCL 3D print have very strong reflections of this kind. In the case of the 2D printouts, it was even only possible at certain angles to capture images at all, where not the complete face is superimposed by this effect. The main difference between two images is the overexposure in some places, apparently due to the material. However, this effect also occurs in the images of the bona fide group, and is therefore not suitable for a targeted evaluation. These observations lead

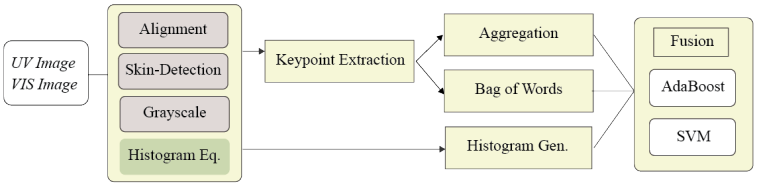

Fig. 2: Flow-Diagram of the proposed Methodology.

us to evaluate these properties in two different ways. If there are no differences between the two images, as would be the case with real skin (MFP), this is an important feature, which can be characteristic of attacks and bona fide. Secondly, there are differences in the ratio of the brightness from VIS to UV, which may differ from those of the skin, they can be seen as spectral signatures. As our database is relatively small, we could not effectively use any deep- or transfer learning approaches. Therefore, we chose conventional features to analyze those characteristics and prove their significance. Since both properties only affect the skin, we use the same preprocessing steps for both methods , which is described in the next section. Our method for extracting the different details of both images are explained in Section 4.2. The brightness differences are presented in Section 4.3.

4.1 Pre-processing

Initially, face detection is performed on the full resolution images. After that, VIS and UV images are aligned to the face region by using face alignment by Zhang et al.[Zh16]. We aligned several images manually in order to guarantee their meaningful inclusion into the dataset. Since it is not expected that the eye region will provide valuable information, we remove this region with a mask. Since hair and the mouth region also contain no valuable information, we perform skin detection by using the procedure of Buza et al. [BAO17] and mask-out all non-skin pixel. In a next step, we convert all images to grayscale, in order to reduce the complexity of our small data-set. In our approach, which evaluates the similarity of local features (See Section 4.2), we also do histogram equalization, which we do not do in the case of brightness analysis (See Section 4.3).

4.2 Analysis of Similarity using local Features

As already shown in previous work[Sa18], MFP features can be extracted effectively via keypoints (KP). We expected to find these properties which are visible in the UV spectrum in high frequency features with a pixel size between 3 and 20 pixels (px). We have therefore selected the ORB (Oriented FAST and Rotated BRIEF) feature detector [Ru11] to extract this property. The ORB detector is computationally very efficient with similar matching performance to SIFT but less affected by image noise and can be used in real-time. A maximum of 1000 ORB KP are calculated using the harris score ranking and four points to produce the oriented BRIEF descriptor. Matching is done by using the euclidean distance between two points, one in the UV image, one in the VIS image, assuming that they denote the same feature if the euclidean distance is smaller than 10 px. In Figure 3 the results of the KP extraction and matching is shown on two images. In the upper images of an attack, with unpainted latex mask, it can be seen that hardly any of them are detected on the surface. It can only be found along the mouth, while in the bona fide image (below) they are recognized throughout all many of them can be matched. The overexposed area on the forehead in the UV image is also clearly visible. With the described method we

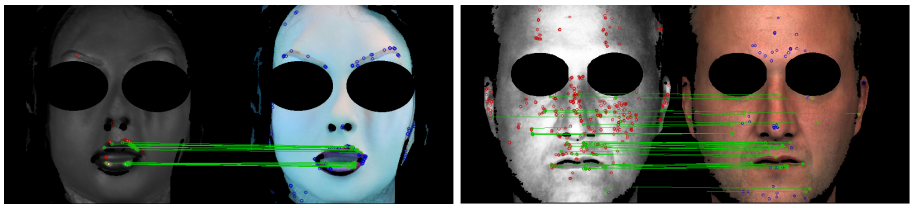

Fig. 3: Matched and Unmatched Keypoints in a typical bona fide image of Skintype II (bottom) and an Attack using a Silicone Mask (top).

have extracted the keypoints on all image pairs. We can detect the following three main differences between attacks and bona fide: (1) The number of detected keypoints is smaller for attacks compared to non-attacks (see Figure 3A)(2) Bona fide show more KP on the UV image that can't be matched with ones on the VIS image (see Figure 3B). (3) In attacks, more unmatched KP can be found on the VIS image than are found on the UV image. The

3D silicone imprint exhibits an extremely high number of unmatched keypoints on both images. These attributes allow us to distinguish both classes in particular, we visualized the number of unmatched KPs in the UV and the VIS image over all classes in Figure 4. We assume that the number of unmatched keypoints between UV and VIS, as well as between VIS and UV contain discriminative information. Therefore, we compose our feature vector as follows: (1) Total KP detected in UV (2) Total KP detected in VIS (3) Matched keypoints (4) Unmatched KP in the UV image and (5) the number of KP in VIS that couldn't be matched to the UV image.

4.3 Analysis of Brightness Property

As can be seen in Figure 3, attacks reflect differently from bona fide faces when captured with an UV camera. Figure 4 (Left) depicts the average difference between VIS and UV images of both bona fide faces and attacks presented in a gray-scale histogram. Thus, this method utilizes the distribution of their brightness values in the form of histograms. It aims to discern legitimate images from attacks by comparing the histograms of both the UV and the VIS image of faces. Since the histograms represent the image's brightness distribution, each has a length of 255. By combining both histograms for one face, we create feature vectors containing the amount of pixels that are of each particular brightness for both the UV and VIS image. We experiment with different methods of combining, including adding, subtracting and concatenating the histograms for feature vectors of a length of either 255 or 510.

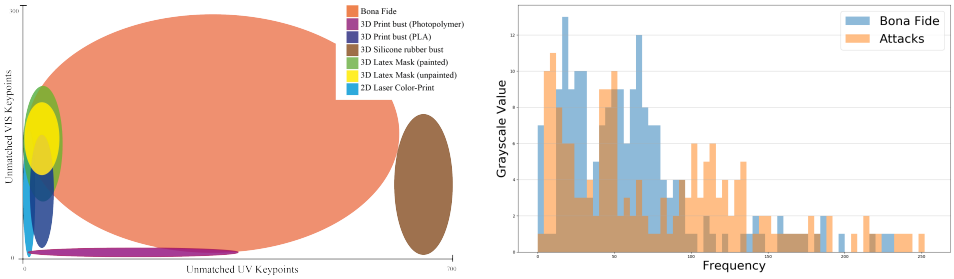

Fig. 4: (Left) Unmatched ORB Keypoints in the UV and VIS Image. (Right) Histogram comparison of Attacks(red) and Bona Fide (blue) in Grayscale.

5 Experiments and Results

In our approach using keypoints, we experimented with adding and omitting the features written in Section 4.2. Here, the variant using all five values has proved to be the best. Using the positions of the keypoints, we experimented with different feature vector lengths between 150 and 500. A length of 300 has proven to be optimal. The feature vectors of the histogram approach are created by either concatenating, adding or subtracting the histograms of UV and VIS photo for a total of 9 experiment setups. The different setups

are then evaluated based on APCER[3] and BPCER[4]. While the SVM and AdaBoost approaches both yielded usable results (with the AdaBoost approach performing the best). The logistic regression approach was not able to capture the difference of legitimate faces and attacks to an acceptable degree. This is likely due to the high amount of data required to train neural networks in comparison to SVM and AdaBoost. Among the different vector combination approaches, adding and concatenating performed comparatively (adding performing slightly better), while subtracting did not perform as well, likely due to a loss of information when brightness value resulted in zero.

Tab. 1: Our Results on the presented Dataset.

Scenario	Histogram		Keypoints		Fused	
	APCER	BPCER	APCER	BPCER	APCER	BPCER
Only Skintype 1-2	0%	0.4%	2.2%	2.45%	0%	0%
Only Skintype 3-4	0%	0.4%	3.3%	3.0%	0%	0%
Only Skintype 5-6	0%	0.4%	3.9%	6.9%	0%	0.2%
All	0.4%	1.2%	4.2%	7.2%	0%	0.2%

Due to the small amount of data available, the evaluation is performed using a leave-one-out approach. Since AdaBoost has showed the best results in all scenarios, we only indicate the error rates using that classifier. We were able to achieve a APCER of 0.4% at 1.2% BPCER for the histogram features. Using this feature, we observed false positives (FP) especially in cases using the 2D print attacks. In case of the KP feature vector we achieved 4.2% APCER at 7.2% BPCER while having FP mostly at the attacks using the silicone 3D print and the painted latex masks. By combining both feature vectors into a common one and training them with AdaBoost we were able to reduce the APCER to 0% at 0.2%. This is consistent with our assumption that both properties contain complementary information that together allow a meaningful distinction of the classes.

6 Conclusion

We presented an experimental study on evaluating the vulnerability of face recognition system towards presentation attacks. We proposed a novel multispectral face image database comprised of 91 subjects and several face presentation attacks. We explored the intrinsic characteristics of UV and VIS images and used global and local features to quantify the captured images as bona fide or attack. Our results indicate that UV images include useful information for PAD.

Acknowledgment

This research work has been funded by the German Federal Ministry of Education and Research and the Hessen State Ministry for Higher Education, Research and the Arts within their joint support of the National Research Center for Applied Cybersecurity ATHENE.

[3] Proportion of attack presentations using the same PAI species incorrectly classified as bona fide presentations in a specific scenario

[4] Proportion of bona fide presentations incorrectly classified as presentation attacks in a specific scenario

References

[ASJT17] Aaron S. Jackson, Adrian Bulat, Vasileios Argyriou; Tzimiropoulos, Georgios: , 3D Face Reconstruction from a Single Image, 2017. http://cvl-demos.cs.nott.ac.uk/vrn/.

[BAO17] Buza, Emir; Akagic, Amila; Omanovic, Samir: Skin detection based on image color segmentation with histogram and K-means clustering. In: 2017 10th International Conference on Electrical and Electronics Engineering (ELECO). IEEE, S. 1181–1186, 2017.

[DD16] Damer, Naser; Dimitrov, Kristiyan: Practical View on Face Presentation Attack Detection. In: BMVC. 2016.

[De12] De Marsico, Maria; Nappi, Michele; Riccio, Daniel; Dugelay, Jean-Luc: Moving face spoofing detection via 3D projective invariants. In: 2012 5th IAPR International Conference on Biometrics (ICB). IEEE, S. 73–78, 2012.

[Fi88] Fitzpatrick TB: The validity and practicality of sun-reactive skin types i through vi. Archives of Dermatology, 124(6):869–871, 1988.

[In16] International Standards Organization: , ISO/IEC 30107-1:2016 - Information technology – Biometric presentation attack detection – Part 1: Framework, 2016.

[Ra17] Raghavendra, R.; Raja, K. B.; Venkatesh, S.; Cheikh, F. A.; Busch, C.: On the vulnerability of extended Multispectral face recognition systems towards presentation attacks. In: 2017 IEEE International Conference on Identity, Security and Behavior Analysis (ISBA). S. 1–8, Feb 2017.

[RB17] Ramachandra, Raghavendra; Busch, Christoph: Presentation attack detection methods for face recognition systems: a comprehensive survey. ACM Computing Surveys (CSUR), 50(1):8, 2017.

[Ru11] Rublee, Ethan; Rabaud, Vincent; Konolige, Kurt; Bradski, Gary R: ORB: An efficient alternative to SIFT or SURF. In: ICCV. Jgg. 11. Citeseer, S. 2, 2011.

[Sa18] Samatzidis, T.; Siegmund, D.; Goedde, M.; Damer, N.; Braun, A.; Kuijper, A.: The Dark Side of the Face: Exploring the Ultraviolet Spectrum for Face Biometrics. In: 2018 International Conference on Biometrics (ICB). S. 182–189, Feb 2018.

[SKJ16] Steiner, H.; Kolb, A.; Jung, N.: Reliable face anti-spoofing using multispectral SWIR imaging. In: 2016 International Conference on Biometrics (ICB). S. 1–8, June 2016.

[Zh16] Zhang, Kaipeng; Zhang, Zhanpeng; Li, Zhifeng; Qiao, Yu: Joint face detection and alignment using multitask cascaded convolutional networks. IEEE Signal Processing Letters, 23(10):1499–1503, 2016.