What you can't see can help you – extended-range imaging for 3D-mask presentation attack detection

Sushil Bhattacharjee
Biometrics Privacy and Security Group
Idiap Research Institute
Martigny, Switzerland
sushil.bhattacharjee@idiap.ch

Sébastien Marcel
Biometrics Privacy and Security Group
Idiap Research Institute
Martigny, Switzerland
sebastien.marcel@idiap.ch

***Abstract*—High-quality custom-made 3D masks are increasing becoming a serious threat to face-recognition systems. This threat is driven, in part, by the falling cost of manufacturing such masks. Research in face presentation-attack detection (PAD) in general, and also specifically for 3D-mask based attacks, has mostly concentrated on imagery in the visible-light range of wavelengths (RGB). We look beyond imagery in the visible-light spectrum to find potentially easier solutions for the challenge of face presentation-attack detection (PAD). In particular, we explore the use of near-infrared (NIR) and thermal imagery to detect print-, replay-, and 3D-mask-attacks. This preliminary study shows that both NIR and thermal imagery can potentially simplify the task of face-PAD.**

***Keywords*—Face Presentation Attack Detection, 3D-Masks, RGB/depth cameras, thermal cameras, NIR, LWIR.**

I. Introduction

Typical face-recognition (FR) methods are highly susceptible to *presentation attacks* (PA), also commonly called *spoof attacks* [5], [11]. The term 'presentation attacks' covers both *impersonation* as well as *concealment* attacks [6]. Most research efforts on presentation attack detection (PAD) so far, have considered specific kinds of attacks, and have proposed solutions to such attacks under specific conditions. Mainly 2D impersonation attacks, that is, attacks performed using 2D presentation-attack instruments (PAI) such as printed paper (*print attacks*) and replay on digital screens (*replay-attacks*) have been studied. The proposed solutions have usually considered the scenario where the biometric sensor is a single color-camera. These solutions are very well-suited for certain applications, such as for current mobile devices. The sensors on such devices are standard color cameras. Therefore, the same sensor-data is typically used for both, face-recognition, as well as for face-PAD. The PAD solutions so far have mainly relied on features extracted from color images and sophisticated machine learning algorithms, to delineate that ever-shrinking margin between the two classes (*bona fide* and attack presentations). One reason why PAD research has evolved in this direction is that adding extra sensors has often been infeasible, due to cost and space constraints.

(a) Custom-made rigid masks

(b) Generic silicone mask

(c) Custom-made silicone mask

Fig. 1: Examples of realistic 3D masks. Approximate prices: (a) US$300, (b) US$800, and (c) US$4000.

PAs using custom-made 3D masks are receiving increasing attention [5], [9]. Figure 1 shows examples of 3D-masks, including rigid masks and soft silicone masks. Highly realistic 3D-masks, such as the custom-made silicone mask shown in Fig. 1(c), are still quite expensive. However, the process of manufacturing custom-made silicone masks is evolving rapidly, and such masks will be available at accessible prices in the near future. The study presented in this paper is motivated by the threat that attacks on FR systems using high-fidelity custom masks may soon become as commonplace as print-attacks and replay-attacks today.

A new category of consumer-grade imaging devices, collectively referred to as *extended-range imaging devices*, are now available that capture data not only in visible-light wavelengths (*i.e.*, color images), but also in near-infrared (NIR) and long-wave infrared (LWIR) domains. Color/depth (RGB-D) cameras, for example, the Microsoft Kinect and Intel RealSense series of products, and thermal cameras can provide easy solutions for once challenging PAD problems.

In this article, we explore the use of such devices to simplify the task of face-PAD for both 2D and 3D attacks. This hardware based approach is especially well-suited for near-real-time scenarios such as border-control applications. The data and code used for the experiments presented in this paper are made freely available on the web[3].

Following a summary of related research (Section II) on the use of extended-range imaging in face biometrics, as well as on PAD for 3D-mask attacks, we provide, in Section III, brief descriptions of the imaging devices that have been used in this work. Experimental results are presented and discussed in Section IV. Finally, Section V gives a summary of this work, along with an outlook on how this work will evolve in future.

II. Related Work

Previous efforts with specialized imaging-sensors for face biometrics applications have been concentrated mainly on the problem of face-recognition. Bhowmick *et al.* [3] have shown that thermal imagery can be used for face-recognition based on facial vein-patterns. Bebis *et al.* [2] explore the fusion of visible and thermal imagery for face-recognition applications. Although they use face-recognition performance as a metric, the main focus of their work is on evaluating different methods for fusing the data from the two modalities.

Lagorio *et al.* [8] use 3D scans from a Vectra 3D camera in a PAD scenario, to detect curved-paper based print presentation-attacks. More recently Raghavendra *et al.* [13] have published a detailed study on the vulnerability of FR systems in extended multispectral imaging domain, involving seven-band imagery covering the visible light and NIR illumination. Their work shows that all the four studied FR approaches are consistently vulnerable in all imaging-bands considered, except for the 930nm (NIR) band. This finding is consistent with previous research [7] showing that the reflectance of human skin drops significantly in a narrow wavelength band centered at 970nm. Steiner *et al.* [14] have demonstrated the use of multispectral SWIR imagery to reliably distinguish human skin from other materials, and have shown that such multispectral devices can be used for PAD.

[3]https://pypi.python.org/pypi/bob.paper.BioSig2017_3DMaskPreStudy

(a) RealSense SR300

(b) Xenics Gobi-640

Fig. 2: Cameras used in this work.

Good quality 3D masks present clear threats in both impersonation as well as concealment categories. Erdogmus *et al.* [5] have published the 3DMAD dataset, using a set of custom-made rigid masks, for experiments in 3D face-PAD. This dataset has also been used by other research groups [1] in other 2D face-PAD experiments. Liu *et al.* [9] have published the more recent HKBU-MARs dataset containing images of 3D-mask based PAs. They have proposed a remote photoplethysmography (rPPG) based approach to detecting 3D-mask PAs. Both works ([5], [9]) use several variants of local binary patterns (LBP) to demonstrate the vulnerability of FR methods to the 3D-mask PAs.

Manjani *et al.* [10] present an observational study into concealment attacks using 3D-masks. They describe PAD experiments based on the SMAD dataset [10], which consists of public-domain videos collected from the World-wide Web.The dataset, however, is relatively small – including only 65 genuine videos and 65 silicon-mask videos. Although observational studies such as this may indicate association between variables (in this case between the true labels of the test videos and the classifier-score), the influence of other confounding variables here cannot be ruled out. To demonstrate the efficacy of a method for detecting 3D-mask based PAs, it is important to design a controlled experiment to highlight exclusively the causal effect of 3D-masks on the resulting classifier-score.

III. Extended-Range Imagery

In this work we have explored the use of two different cameras – the RealSense SR300 camera from Intel, and the

Gobi-640-GigE LWIR camera produced by Xenics[4] – for PAD applications.

The RealSense SR300 camera (Fig. 2(a)) is Intel's second generation depth-sensing camera. It uses a structured-light approach, based on 860nm NIR illumination, to capture depth information from a 3D surface. This camera has a relatively shallow field of view, and produces the most accurate results in the depth range of 0.2m – 1.2m. In practice, the implication for PAD applications is that the average intensity of the NIR image drops rapidly with distance. Therefore, to capture good quality images, the subject should be positioned quite close (0.3m – 0.5m) to the camera. Besides depth-information, the camera also captures color (RGB) images and NIR images. It is important to note that the two cameras (color and NIR) have different fields of view, and are not mutually calibrated. For the experiments discussed in this paper, we have used both color and NIR images at VGA resolution.

The Xenics Gobi thermal camera, shown in Fig. 2(b), covers a wavelength range of 800nm – 1200nm, and captures 16-bit images at VGA (640×480 pixels) resolution. Although the camera can take a range of lenses, we have used the standard 18mm $f/1$ lens, with a horizontal field of view of 33°.

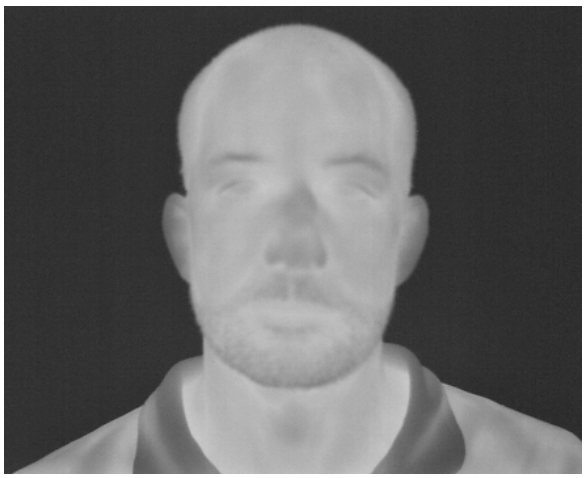
For both cameras, we have developed data-capture tools in-house, based on software development kits (SDK) are available for each camera. When using images outside the visible-light range of wavelengths, one practical problem that arises is that of face-detection. Most face-detection tools take a machine-learning based approach, and need to be trained with sufficient amounts of training-data. Since faces present different spectral characteristics in different wavelengths, these appearance-based face-detection schemes need to be explicitly trained for each imaging modality. We simplify the face-detection process, for each camera, by positioning the subject such that the face falls within a pre-defined rectangle (displayed on the live-display monitor). The position and size of the rectangle can be adjusted for each subject, and is recorded along with the images. Thus, for each camera, face-position information is available directly from the data-capture process.

Sample images of *bona fide* presentations are shown in Fig. 3. The images in Fig. 3(a) and (b) have been captured using the SR300 RGB-D camera, and show the *bona fide* presentation in visible wavelengths (RGB) and NIR band respectively. Figure 3(c) shows an image captured using the Xenics Gobi thermal camera, illustrating the appearance of a *bona fide* presentation in the LWIR band.

[4]Website: www.xenics.com

(a) Visible (RGB)

(b) NIR



(c) Thermal (LWIR)

Fig. 3: Examples of *bona fide* presentation images, as seen in different wavelength bands. Images in (a) visible (RGB) and (b) NIR wavelength-bands have been captured using the SR300. The LWIR band image (c) has been captured using the Xenics Gobi camera.

IV. Experiments

In this section we present experimental results for PAD based on extended-range imagery. Based on some initial tests, we concluded that NIR images have the potential to easily detect 2D and 3D PAs. In the experiments reported here, we have specifically investigated the following questions. (1) Can NIR imagery be useful in detecting 2D PAs? (2) Is it possible to detect 3D-mask attacks (of both rigid and flexible varieties) in NIR images? (3) Can we use thermal (LWIR) images to detect custom-made flexible-mask PAs?

A. PAD Using NIR Imagery

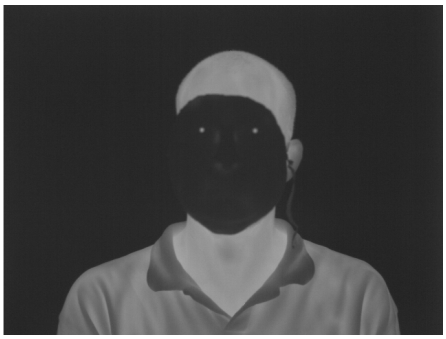
We start by confirming the intuition that NIR imagery can simplify the task of detecting 2D PAs such as print- and replay-attacks. Figure 4 illustrates typical images captured by the SR300 camera for various kinds of PAs. Fig. 4(a) and (d) show the appearance of a print-attack in color- and NIR-imagery, respectively. We note that although the face may be detected in the color image (Fig. 4(a)), even a simple image-histogram analysis would be sufficient to determine that no face present in the corresponding NIR image (Fig. 4(d)). The analysis for the digital-replay attack shown in Fig. 4(b) and (e) is analogous.

Fig. 4(c) and (f) show the color and NIR images for the same rigid mask attack. Visual inspection of the mask region shows that the simulated facial features, such as the painted eye-brows and moustache, are entirely suppressed in the NIR image, and the mask region has a texture-less appearance. Intuition tells us that detecting such surfaces in NIR images should be reasonably easy. Contrary to initial expectations, however, detecting such 3D mask PAs in NIR images is not straightforward. Comparing Fig. 4(f) with Fig. 3(b), we see that the NIR image presents similar image characteristics for both *bona fide* and 3D-mask attack presentations. Indeed, our preliminary tests with NIR-860nm band images showed that lower-order statistics (intensity-histograms, histograms of oriented gradients (HOG) and gray-level co-occurence matrices (GLCM)) of the 3D-mask presentations are quite similar to those of *bona fide* presentations. Detection of 3D mask PAs in NIR images would require more complex processing such as, modeling of the peri-occular region.

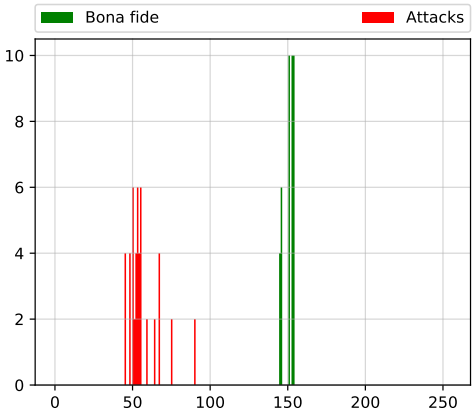
Although counter-intuitive, this result is quite logical. The NIR wavelength used by the RGB-D cameras has been deliberately chosen to be such that the illumination is strongly reflected by most kinds of surfaces, including human-skin. This is imperative for the primary purpose of the camera – capturing depth-information using structured-NIR illumination.

B. PAD Using Thermal Imagery

(a) *Bona fide*



(b) 3D-mask attack



(c) Intensity distribution

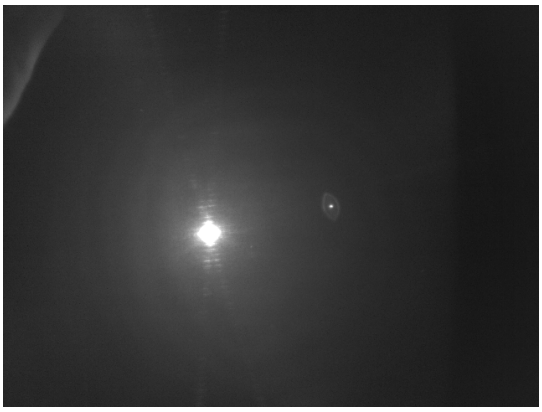
Fig. 5: Examples of thermal images of (a) *bona fide* and (b) 3D-mask attack presentations. (c) Histograms of average-intensity over the face-region, for mask- and *bona fide* presentations, computed over a small dataset of thermal images.

Thermal (LWIR) images are very well suited for detecting 3D-masks PAs with high certainty. Figure 5 shows images from the Xenics Gobi thermal camera. The mask in Fig. 5(b), being cooler than the body-temperature of the subject, is clearly demarcated. Figure 5(c) shows distributions of the average pixel-intensity of a small region centered on the face. This plot has been generated based on a small dataset of *bona fide* and 3D-mask attack presentations using five subjects and

(a) Print; Color

(b) Print; NIR

(c) iPad; NIR



(d) iPad; Color

(e) 3D Mask; Color

(f) 3D Mask; NIR

Fig. 4: Various PAs, as seen by the RGB-D (SR300) camera. Three kinds of attacks – print, replay, and 3D-mask – are shown. Left column (a), (c), (e): the appearance of print, replay, and 3D-mask attack, respectively, in visible color wavelengths. Right column (b), (d), (f): corresponding images of the respective scenes under NIR illumination.

six rigid masks. The intensity distribution for the rigid-mask presentations is plotted in red, and the distribution for the *bona fide* presentations is plotted in green. Although this plot is based on a very small dataset, it indicates that the average intensity over the face-region is significantly lower when a rigid-mask is used.

Flexible silicone masks are often hand-finished, and offer greater color and texture fidelity, and therefore, pose a greater threat, compared to rigid masks. The images in Fig. 6 allow us to compare the appearances of *bona fide* presentations and flexible custom-made silicone-mask attack presentations, under visible-light and NIR illuminations. The mask shown here (Fig. 6(c), (d)), which is the same as the mask shown in Fig. 1(c), has been custom-designed to match the face of the subject shown in Fig. 6(a) and (b)[5]. Although the mask may be easily apparent to the human observer, FR systems are quite vulnerable to such mask-attacks.

(a) *Bona fide*; RGB

(b) *Bona fide*; NIR

(c) Flex. mask; RGB

(d) Flexible mask; NIR

Fig. 6: Comparison of *bona fide* and flexible silicone mask attack presentations in visible-light and NIR illuminations. Images captured using the SR300 camera.

Table I shows the face-recognition scores for the various attack-attempts on the reference-subject shown in Fig. 6(a)&(c). The pre-trained VGG-Face neural network [12] has been used for face-recognition in this experiment. Specifically, the table shows scores for five zero-effort impostor (ZEI) presentations, one genuine presentation, and one attack presentation made using a custom-made silicone-mask. Scores are shown for two image-modalities – color images and NIR images. The first row (labeled 'RGB') shows scores of the various presentations using color images, where the target identity is the image in Fig. 6(a). Scores for a similar experiment using NIR images, with the image in Fig. 6(b) as the target identity, are shown in the last row of the table. In both experiments we can see that the score for the genuine presentation is at least an order of magnitude higher than the scores for the ZEI presentations (Subjects 1-5). It is interesting to note that, in both illumination wavelength bands, the score for the custom-mask attack is much closer to that of the genuine presentation, than to the ZEI presentations. This small-scale experiment cannot be attributed any statistical significance. It does, however, emphasize the necessity for a large-scale study involving attacks with high-quality custom-made silicone masks.

[5]The mask was manufactured based on 3D facial scans of the subject, using the Intel RealSense F200 camera, which is the predecessor of the SR300 camera. Additional color photographs were used to ensure high-quality finish for the visual appearance of the mask.

C. Discussion

Preliminary experimental results discussed above show that extended-range imaging devices can drastically simplify the task of face-PAD. The task of face-PAD using such devices, however, in not entirely straightforward. For example, the main reason why the face in the print-attack is not visible in 4(d)) is that the ink used here is not IR-reflective[6]. There are, however, IR-reflective inks available on the market, that will show a strong response in NIR-band images [4]. Further experimentation is therefore necessary to tackle the challenge print-attacks constructed using IR-reflective inks.

In Fig. 5 we have seen how rigid 3D-mask attacks can be easily detected using thermal cameras, because the masks have a much lower temperature than average human body temperature. This advantage is lost to some extent, when dealing with flexible silicone masks, which can warm up very quickly when in contact with warmer objects, such as an attacker's face. Figure 7 illustrates the evolution of the temperature of a silicone mask when worn by a human subject. The figure shows six frames of a time-lapse sequence (captured using the Xenics Gobi camera), with an interval of 30 seconds between frames. From left to right, the change in temperature of the mask is evident, especially in the forehead region, where the mask makes good contact with the subject's face. The top of the mask can be clearly distinguished in the first frame on the left. In the right-most frame, this region seems to be almost as warm as the subject.

There is a clear need for research into PAD of flexible 3D masks, where the limits of thermal imagery for detecting custom-made masks is explored under various conditions, using a large set of silicone masks. Thermal cameras such as the Xenics Gobi used in this work are still quite expensive. However, low-cost options are now becoming increasingly available. Cameras such as the FLIR-One and the Seek-Thermal Compact Pro are designed to work with mobile phones, offer adequate spatial resolution, and are available at relatively modest prices.

V. Conclusions

Due to the affordability of modern RGB-D cameras, researchers are beginning to look into such cameras for face-recognition applications. These cameras typically use NIR illumination to recover depth information from the scene. Some

[6]Note that the specific print-attack used in this example has been printed according to the specifications prescribed in the Norwegian project SWAN.

	Subject 1	Subject 2	Subject 3	Subject 4	Subject 5	Genuine	Mask-attack
RGB	-0.344	-0.253	-0.265	-0.293	-0.263	-0.038	-0.163
NIR	-0.330	-0.227	-0.280	-0.216	-0.198	-0.028	-0.096

Tab. I: Face recognition scores using VGG network.

Fig. 7: Silicone masks can warm up very quickly when worn.

RGB-D cameras also return 2D videos in visible-light and NIR wavelengths. In this paper we have presented a preliminary study into the utility of NIR and LWIR imagery for face-PAD. Our tests with Intel's RealSense SR-300 camera show that images in NIR wavelength-band can be used to easily detect various 2D presentation-attacks, such replay-attacks and certain kinds of print-attacks. Some kinds of printer-ink do show a strong response under NIR illumination at certain wavelengths. In future work we will investigate methods for detecting print-attacks created using such inks. Contrary to initial expectations, however, monochromatic NIR imagery, of the kind provided by low-cost RGB-D cameras, may not be effective for straightforward detection of 3D-mask attacks.

Realistic custom-made silicone masks will soon be available at affordable prices. Tt is, therefore, imperative to develop face-PAD methods that are robust to 3D-mask based attacks. The examples presented here indicate that the use of NIR and LWIR imagery for detecting rigid as well as flexible 3D masks seem to be promising research directions. We would like to invite the entire biometrics research community to pool resources to create and share a large and diverse data-set for the purposes of such research.

Thermal cameras have been available for several decades. Until recently, however, they have been very expensive, and have not been considered for PAD applications for reasons of cost. Low-cost thermal cameras, such as the FLIR-One[7] and the Seek-Thermal Compact Pro[8], have recently appeared on the market. These devices are designed to work with mobile phones. They offer adequate spatial resolution, and are available at reasonable prices. In future experiments we plan to explore the applicability of such cameras for PAD.

ACKNOWLEDGEMENT

This work has been supported by the EU H2020 project TeSLA, the Norwegian project SWAN, and the Swiss Center for Biometric Research and Testing. We gratefully acknowledge the critical help from our colleague, Mr. Guillaume Clivaz, who implemented the image-capture applications for the two cameras used in this work.

[7]Website: www.flir.com/flirone

[8]Website: www.thermal.com/compact-series.html

REFERENCES

[1] A. Agarwal et al. Face Anti-Spooïng using Haralick Features. In *Proc. IEEE Intl. Conf. BTAS*, Niagara Falls, NY, USA, 2016.

[2] G. Bebis et al. Face Recognition by Fusing Thermal Infrared and Visible Imagery. *Image and Vision Computing*, 24(7):727–742, July 2006.

[3] M. Bhowmick et al. *Thermal Infrared Face Recognition - a Biometric Identification Technique for Robust Security System*, pages 135–162. Intech, Rijeka, Croatia, 2011.

[4] I. Chingovska et al. *Face recognition in extended imaging domain*, pages 165–194. Springer, 2015.

[5] N. Erdogmus and S. Marcel. Spooïng in 2d Face Recognition with 3DMasks and Anti-spooïng with Kinect. In *Proc. IEEE Intl. Conf. BTAS*, Washington D.C., 2013.

[6] ISO/IEC DIS 30107-1. information technology – Biometric presentation attack detection – Part 1: Framework. Iso standard, Geneva, CH, Jan. 2016.

[7] Y. Kanzawa et al. Human Skin Detection by Visible and Near-Infrared Imaging. In *Proc. IAPR Conf. on Machine Vision Applications (MVA2011)* , Nara, Japan, 2011.

[8] A. Lagorio et al. Liveness Detection based on 3D Face Shape Analysis. In *Proc. IEEE Intl Workshop on Biometrics and Forensics (IWBF)*, Lisbon, Portugal, 2013.

[9] S. Liu et al. A 3D Mask Face Anti-spoofing Database with Real World Variations. In *Proc. IEEE Conf. on Comp. Vision and Patt. Rec. Workshop (CVPRW)*, Las Vegas, 2016.

[10] I. Manjani et al. Detecting Silicone Mask-Based Presentation Attack via Deep Dictionary Learning. *IEEE Trans. Info. Forensics and Security*, 12(7):1713 – 1723, 2017.

[11] S. Marcel, M. S. Nixon, and S. Z. Li, editors. *Handbook of Biometric Anti-Spoofing*. Springer-Verlag, 2014.

[12] O. M. Parkhi et al. Deep face recognition. In *British Machine Vision Conference*, 2015.

[13] R. Raghavendra et al. On the Vulnerability of Extended Multispectral Face Recognition Systems Towards Presentation Attacks. In *Proc. IEEE Intl. Conf. on Identity, Security and Behavior Analysis (ISBA)*, New Delhi, 2017.

[14] H. Steiner et al. Design of an active multispectral SWIR camera system for skin detection and face verification. *Journal of Sensors*, (1):1 – 8, 2016. Article ID 9682453, Special Issue on Multispectral, Hyperspectral, and Polarimetric Imaging Technology.