Indirect synthetic attack on thermal face biometric systems via visible-to-thermal spectrum conversion

Khawla Mallat and Jean-Luc Dugelay
EURECOM
Sophia-Antipolis, France
{mallat, dugelay}@eurecom.fr

Abstract

The growing interest in exploring thermal face biometrics is mostly due to the robustness of thermal imagery to face spoofing attacks. However, this robustness lies in the acquisition of thermal properties by the thermal sensor and is limited to presentation attacks. In this paper, we propose a new type of attack on thermal face recognition systems, performed at the post-sensor level. In the visible spectrum, this attack would be carried out by simply injecting a face image of the claimed identity into the communication channel right after the sensor. However, unlike visible face images that are abundantly available on the web, thermal face images are not easy to obtain. Therefore, we propose to generate synthetic thermal attacks by converting visible face images into the thermal spectrum. To perform visible-to-thermal spectrum conversion, we use a cascaded refinement network trained using contextual loss. In a scenario where the attacker has prior knowledge about the spoofing countermeasure of the system, we introduce a new loss computed at the local binary pattern (LBP) maps level to fool an LBP-based spoofing attack detection algorithm. The vulnerability of thermal face biometric systems to the proposed attack is then assessed using two existing baselines of spoofing attack detection. When compared to the challenging presentation attack using silicone masks, the equal error rate has increased from 0.20% to 11.60% and from 2.28% to 58.54% when exposed to the proposed synthetic attack, using the two spoofing attack detection baselines.

1. Introduction

Automatic face recognition is one of the most intuitive and convenient biometric systems as it offers a contactless and non-intrusive process. Face biometric systems have immensely evolved to achieve human-level performance. For these reasons, face biometric systems are now widely deployed in countless applications from border control to face unlock on mobile devices. With this growing usage of face biometric systems, it is commonly acknowledged that this technology is exposed to multiple threats [12, 18, 21]. Eight different levels of attacks have been defined in [12, 24], a ninth level of attack that occurs on the spoofing countermeasure unit can be considered. Figure 1 illustrates the different levels of attacks on face biometric systems. The two most vulnerable points in a face biometric system are marked by 1 and 2 in Figure 1, corresponding to *direct* or *physical access* and *indirect* or *logical access*. Face biometric systems might be the most vulnerable among all biometric systems, as faces are accessible on social networks or through capturing a photograph at a distance without the victim's consent.

Direct or physical access attacks occur at the pre-sensor level and is referred to as presentation attack. According to *ISO/IEC30107* standards [2], presentation attack is defined as *"the presentation of an artefact or of human characteristics to a biometric capture subsystem in a fashion intended to interfere with system policy"*. This attack can be carried out either to impersonate/spoof a genuine user to gain unauthorised access, or to evade the biometric system by concealing the attacker's identity. The presented artefact can consist of a fake biometric sample of the claimed identity (photograph, mask, etc) in spoofing scenarios, or some alteration or falsification [8, 9] applied to the attacker's biometric sample in evasion scenarios. Recent research studies [3, 11, 5, 8, 23] have proved that using thermal imagery might be the most effective solution to presentation attack detection. Thermal imagery detects electromagnetic radiations in the mediumwave MWIR (3 - 8μm) and longwave infrared spectrum LWIR (8 - 15μm) in which most of the heat energy of human faces is emitted. The thermal signature of the human face thus provides evidence of the user's liveness. Artifacts presented by the impostor exhibit different thermal characteristics of those of a face, leading to a straightforward presentation attack detection solution.

Indirect or logical access attacks, on the other hand, occur at the post-sensor level. For this scenario, it is assumed that the attacker has access to the communication channel

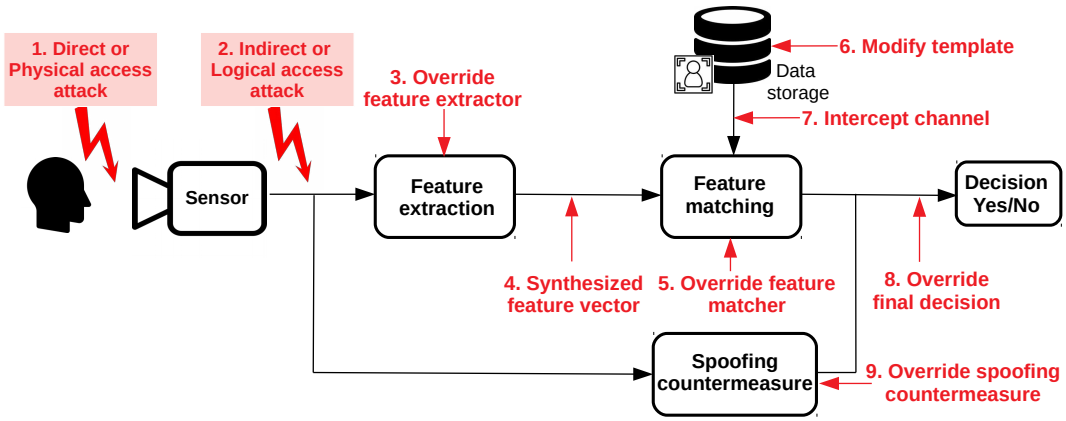


Figure 1: Levels of attacks on face biometric systems.

between the sensor and the feature extraction module, as shown in Figure 1. This attack intercepts the face sample acquired by the sensor and substitutes it with a fake sample of the claimed identity. These attacks in the visible spectrum include replay attacks [12], adversarial attacks [26], face morphing attacks [20]. Face samples are easy to obtain to spoof conventional visible spectrum-based face biometric systems. However, this isn't the case for thermal face biometric systems, as thermal images are not abundantly available as is the case of visible images.

While until very recently the deployment of thermal technologies would have been very expensive to deploy and thus an un-realistic alternative to presentation attack detection, the use of thermal imagery is now a reality. It is perhaps for this reason thermal imagery is gaining a lot of attention and starting to be deployed across many applications requiring high levels of security. Therefore, it is essential to study all the vulnerabilities of thermal face biometric systems and the threats they may encounter.

In this paper, we propose a new attack on biometric samples at the post-sensor level for thermal face biometric systems. The robustness of thermal face biometric systems lies in the process of acquisition characterising thermal sensors. Therefore, the indirect access attacks, that occur at the post-sensor level, are an obvious threat to the reliability of thermal face biometric systems. We presume that the attacker injects into the thermal face biometric system a fake thermal face sample representing the thermal signature of the claimed identity. This type of attack on thermal face biometric systems, to the best of our knowledge, has not yet been explored in the literature. Since thermal face images are nearly impossible to obtain, the new attack we are proposing consists of generating synthetic thermal face images by converting images acquired in the visible spectrum to the thermal spectrum. We also consider the scenario where the attacker has prior knowledge about the spoofing countermeasure used in the system and uses it to adapt his/her synthetic attack to better spoof the system. We assess the vulnerability of the existing countermeasure approaches designed for thermal spectrum when exposed to the proposed synthetic attacks.

The remainder of this paper is organised as follows. Section 2 presents briefly the studies carried out for spoofing attacks on the thermal spectrum. Section 3 introduces our approach to generate the proposed thermal attack, and the modifications we applied to obtain a more challenging attack for a given spoofing attack detection approach. Section 4 details the process to generate the proposed synthetic attacks and presents a quality assessment of the generated thermal images. Section 5 reports the experimental setup defined for the evaluation of two existing baselines of spoofing attack detection when exposed to the proposed attacks, followed by results and discussion. Conclusions are presented in Section 6.

2. Related work

First attempts of spoofing attacks included techniques as simple as the presentation of a photograph from the claimed identity on a printed paper or on a mobile device screen, which can alter the performance of algorithms operating exclusively on 2D images. Some prompt solutions have been proposed such as requiring an eye blink, smile, or other visual reactions to prove the liveness of the user, yet this can be easily tricked using video replay attacks. New sensor-based presentation attack countermeasures have also been considered, as these sensors deliver complementary visual information. 3D sensors [7, 13] unravel the lack of depth information when a printed photograph or a video played on a device is presented. A much more robust sensor against these attacks is thermal cameras, as it provides proof of the user's liveness simply through acquisition [4]. When

presenting these aforementioned attacks, the acquired thermal sample will present some properties that are different from those of a human face thermal signature. More elaborate and high-cost methods of spoofing have later appeared to manufacture 3D masks, which are robust to 3D sensor-based presentation attack detection. Thermal sensors remain highly robust against rigid 3D mask attacks, as the rigid mask presents a uniform pattern with a much lower temperature than an average human face. However, this robustness can be affected when a flexible silicone or latex mask-based attack is presented, as it can get heated when worn by the attacker's face. Recent studies [3, 11, 5] do however show that even though the robustness of thermal sensor-based presentation attack detection drops, thermal modality remains the most robust among other studied modalities such as visible spectrum, depth maps, and near-infrared spectrum. As for evasion, the attack can consist of face disguise and it can go as far as getting plastic surgery. While this can practically interfere with visible spectrum-based face biometric systems, thermal technology has been proved substantially robust to these attacks as well [8, 23]. Face disguise can easily be detected since the used accessories present different thermal properties from those of a human face [8]. Thermal imagery can also identify plastic surgeries, as the resulted alteration of blood vessels appear as cold areas in the face [23].

A preliminary study was carried out, by Bhattacharjee *et al*. [4], to explore the usage of multi-channel information for presentation attack detection. The study considered, along with the visible spectrum, data from thermal, near-infrared, and depth channels. The authors demonstrated that 3D masks and 2D attacks can easily be detected in the thermal spectrum by using the mean facial intensity of the face region. In [5], the authors proved the vulnerability of commercial face recognition systems to custom silicone masks. They also propose, as a solution for presentation attack detection, to use the mean facial intensity, as proposed in [4]. Agarwal *et al*. [3] introduced a multispectral database of latex mask attacks including visible, near-infrared, and thermal spectra. The authors performed different experiments for face verification and presentation attack detection independently for each spectrum. For presentation attack detection, they proved that the thermal spectrum is the most robust in comparison to visible and near-infrared spectra. The best performing system was based on redundant discrete wavelet transform (RDWT), Haralick features, and support vector machine (SVM). However, the results reported on the thermal spectrum are questionable since the thermal data seems to be acquired using FLIR MSX[1] mode which adds visible light details to the thermal images. George *et al*. [11] present a new multi-channel database containing different 2D and 3D attacks. A multi-channel convolutional neural network was proposed in [11] for presentation attack detection. Besides, a score level fusion was performed combining the scores of each channel's presentation attack detection algorithm. For thermal spectrum, a presentation attack detection algorithm, based on local binary pattern (LBP) feature extraction followed by logistic regression classification, had outperformed the RDWT-Haralick-SVM baseline proposed by [3]. In [8], a disguise database in the visible and thermal spectrum was proposed. The authors proposed to combine patches from visible and thermal images for presentation attack detection.

[1] https://www.flir.com/discover/professional-tools/what-is-msx/

3. Visible-to-thermal spectrum conversion

A new attack on thermal face biometric systems is proposed in this work. This attack occurs at the post-sensor level and is obtained by converting available visible face images to thermal spectrum. In this section, we introduce the used approach to convert visible images to thermal spectrum. A customisation of the used approach is later presented to generate more challenging attacks to a given approach of the thermal spectrum-based presentation attack detection. Finally, implementation details of the proposed approaches are given.

3.1. Cascaded refinement network for spectrum conversion

Visible-to-thermal spectrum conversion was carried out using cascaded refinement networks (CRN) [15, 17]. Selecting CRN network for our visible-to-thermal spectrum conversion was driven by the fact that it scales seamlessly to high-resolution yielding more realistic images. CRN [6] is a feed-forward convolutional neural network consisting of a cascade of refinement modules, each module operating at a given resolution, as illustrated in Figure 2. CRN model synthesises images progressively starting at the lowest resolution (4×4 in our implementation), then feature maps are incrementally refined duplicating the resolution at each step until reaching the target image resolution (128×128 in our implementation).

The parameters optimisation of the CRN network is performed using contextual loss (*CX*) [19], which aims to compare regions with similar semantic details while preserving the context of the entire image. Our loss function can be modeled as a combination of two losses: style loss and content loss, as defined by Gatys et. al [10]. The style loss is computed between the generated thermal image and the ground truth thermal image. Minimising the style loss yields to generate artificial images with the same properties as the target thermal image. The content loss is computed between the input visible image and the generated thermal image. The content loss aims at preserving details as facial attributes, while tolerating some local deformations that are required to perform the visible-to-thermal style conversion.

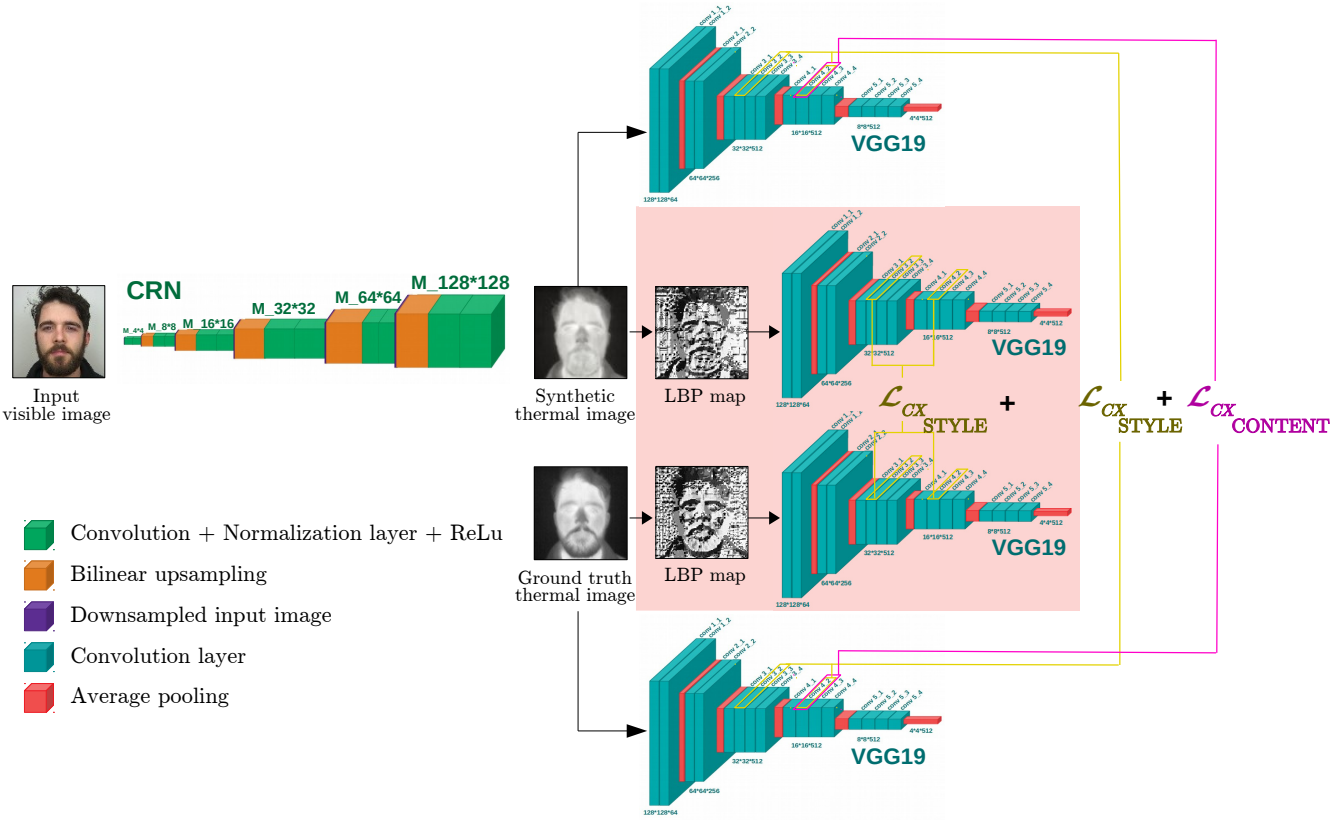


Figure 2: Illustration of the proposed architecture to perform visible-to-thermal spectrum conversion. The highlighted blocks of the diagram illustrate the introduced loss for the customised approach.

Both losses are computed at embedding level, extracted using VGG19 [25]. The CRN loss can be formulated as follow:

$$\mathcal{L}_{CRN}(G, I_{Vis}, I_{Th}) = \alpha_1 \mathcal{L}_{CX}(\Phi_{l_s}(G(I_{Vis})), \Phi_{l_s}(I_{Th})) + \alpha_2 \mathcal{L}_{CX}(\Phi_{l_c}(G(I_{Vis})), \Phi_{l_c}(I_{Vis})) \quad (1)$$

Where I_{Vis}, I_{Th} and G denote the input visible image, the ground truth thermal image and the generator (i.e. visible to thermal synthesis model), respectively. Φ_{l_c} and Φ_{l_s} refer to the VGG19 embeddings extracted at content layers `conv4_2` level and style layers level `conv3_2` and `conv4_2`, respectively.

3.2. Customised CRN for attack synthesis

Here, we explore the scenario in which the attacker has obtained some prior information about the spoofing attack detection approach used in the targeted thermal face biometric system.

The study, carried out by George *et al.* in [11], has proven that the spoofing attack detection algorithm based on LBP feature extraction is outperforming the solution proposed by [3]. Therefore, we consider the LBP-based spoofing attack detection as our reference approach on which the attacker has some prior information. Consequently, we customised our visible-to-thermal spectrum conversion model in a way that it intends to generate thermal images of which the LBP map is more similar to the LBP map of thermal ground truth images, or simply put more similar to the LBP map of thermal bonafide samples. LBP was originally introduced by Ojala *et al.* [22] for texture analysis, but later on, it was extensively explored in numerous applications. Particularly, it has shown its efficiency for face analysis.

In Figure 2, the architecture of the two used approaches to perform visible-to-thermal spectrum conversion is illustrated. Our LBP-based customisation of the CRN model is highlighted in red. The part of the diagram that is not highlighted, represents the basic CRN model, where we observe the loss at content level computed between the input visible image and the synthetic thermal image, and the loss at style level between the synthetic thermal image and the thermal ground truth (bonafide) image. In addition to the loss defined for the basic CRN model, we introduced a new loss

that is computed at the LBP map level. The LBP map is generated using a uniform pattern, 8 sample points in the neighborhood on the circle of radius 1. To compute this loss function, we propose to use contextual loss computed on LBP maps, but solely at the style level, as our objective is to generate thermal attacks of which the LBP maps are closer to the LBP map of thermal bonafide images. We extracted the VGG19 embedding vectors from LBP maps of the synthetic thermal image and the thermal ground truth, at style layers. Consequently, the total loss of the conversion model, in this case, is defined as follow:

$$\mathcal{L}_{Total}(G, I_{Vis}, I_{Th}) = \lambda_1 \mathcal{L}_{CRN}(G, I_{Vis}, I_{Th}) + \lambda_2 \mathcal{L}_{CX}(\Phi_{l_s}(LBP(G(I_{Vis}))), \Phi_{l_s}(LBP(I_{Th}))) \quad (2)$$

In addition to the annotation defined in the equation 1, LBP denotes the LBP map. For the remainder of the paper, we refer to the visible-to-thermal conversion models as *CRN*, *CRN+CX(LBP)* to denote the basic CRN model and the CRN model combined with the contextual loss at style level computed on LBP maps, respectively.

3.3. Implementation details

VIS-TH face database [16] is used to train our visible-to-thermal spectrum conversion models. This database is publicly available and contains face images in both visible and thermal spectra with a pixel resolution of 1920×1080 for the visible images and 160×120 for the thermal images with a spectral response range of 7.5 - 13.5μm. Unlike the few existing databases of visible and thermal face, this database is acquired simultaneously using the dual-sensor camera Flir DUO R [1] considering a wide range of facial variations. The database contains in total 1050 pairs of images collected from 50 subjects of different ages, gender, and ethnicities. During training, we excluded one variation as it was acquired in total darkness, which yields 1000 pairs of face images. Visible and thermal images are registered and re-sampled to 128×128 pixels.

The training of the two proposed models of visible-to-thermal spectrum conversion was performed with a learning rate of 1e-4. The *CRN* model was trained for 40 epochs and *CRN+CX(LBP)* model for 90 epochs. The weights assigned to the different losses α_1, α_2, λ_1 and λ_2 were adjusted using Grid Search.

4. Indirect attack synthesis

In this section, the dataset, from which the synthetic thermal attacks are generated, is first introduced. A quality assessment of the synthetic thermal images is then performed.

4.1. CSMAD dataset for indirect attack synthesis

Choosing the Custom Silicone Mask Attack Dataset (CSMAD) [5] is motivated by the fact that this dataset contains the most challenging attack on thermal face biometric systems, and therefore it will be considered as a baseline attack. In other words, the vulnerability of the spoofing attack detection approaches to the new attack, proposed in this paper, will be assessed and compared to the vulnerability to the silicone masks attack.

CSMAD contains presentation attacks made of six custom-made silicone masks. Face images are collected from 14 subjects. Bonafide samples were collected from all subjects. Extra bonafide samples were acquired for few subjects for which they were wearing eyeglasses. Attack samples were acquired for all 6 masks but worn by different attackers. Additional attack samples were recorded with the masks attached to their provided stands. The CSMAD provides bonafide and attack acquisitions, consisting of videos of 5 to 10 seconds, in the visible, near-infrared and thermal spectrum, and also depth maps collected simultaneously. The dataset was collected under 4 different illumination conditions. In our study, we have only considered data from the visible and thermal spectrum.

Figure 3 presents few attack samples. We can observe, in column (a), where the mask is worn by the attacker it gets warm exhibiting a thermal face sample that looks more like a face. Whereas for the attacks where the mask is attached to a stand, we can barely differentiate the mask from the background in the thermal spectrum, as they probably have similar temperatures.

Figure 3: Samples of presentation attack of CSMAD database in visible and thermal spectrum. **(a)** worn masks **(b)** standing masks.

4.2. Quality assessment of the synthetic attacks

Bonafide samples of the CSMAD dataset, which are acquired in the visible spectrum, are fed to the visible-to-thermal spectrum conversion models presented, in section 3, to generate the synthetic attack. Two of the illumination conditions were discarded as they altered the quality of the synthetic images resulting in black areas in the face caused by missing information in the visible spectrum.

Figure 4 illustrates the synthetic attacks in column (c), and (d). We note that the synthetic thermal images present a realistic pattern of thermal signature. Some details, such as hair and eyebrows, are converted into low pixel values reflecting regions with lower temperature compared to the face region. However, we can observe that the synthetic thermal images, when compared to thermal ground truth in column (b), present more details in some facial traits such as eyes and mouth. This is expected as the synthetic thermal images are generated from data with a different source of information. Comparing the synthetic thermal images generated using the two proposed visible-to-thermal spectrum conversion models, we note that the two sets, (c) and (d), of synthetic images are remarkably similar, we can note few minor differences that are almost not visually perceptible.

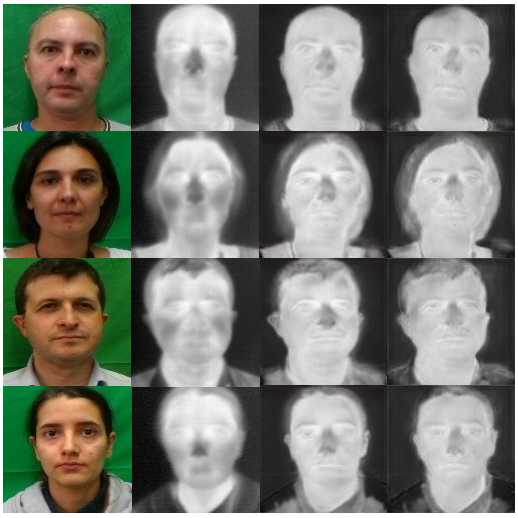


Figure 4: Samples of synthetic attacks. **(a)** visible bonafide **(b)** thermal bonafide (ground truth) **(c)** synthetic attacks using *CRN* **(d)** synthetic attacks using *CRN+CX(LBP)*.

A quality assessment of the synthetic thermal attacks obtained by the two proposed approaches is performed in terms of peak signal-to-noise ratio (PSNR), structural similarity index measure (SSIM), brightness and contrast. PSNR and SSIM are computed between the synthetic thermal images and the thermal bonafide samples (ground truth). The brightness and the contrast quality metrics [14] are measured for each set of images, then a similarity score was deduced by comparing the quality measures of the synthesised thermal images obtained using the two different models to the quality measures of the bonafide thermal images. Table 1 reports the quality metrics obtained for the two visible-to-thermal conversion models. Except for the brightness metric, we observe that the obtained results do not reflect a high fidelity of the synthetic thermal images to the ground truth. As illustrated out in Figure 4, the synthetic attacks are generated from visible face images which provide different information compared to the thermal spectrum. The visible-to-thermal conversion models aim to generate thermal-like images but they cannot predict accurately the thermal signature. However, we observe that the *CRN+CX(LBP)* model generates thermal images that are more similar to the ground truth (bonafide) thermal images than the *CRN* model, which demonstrates the utility of the LBP loss term in improving the quality of the synthesised thermal images.

5. Evaluation of face spoofing attack detection for indirect synthetic attack

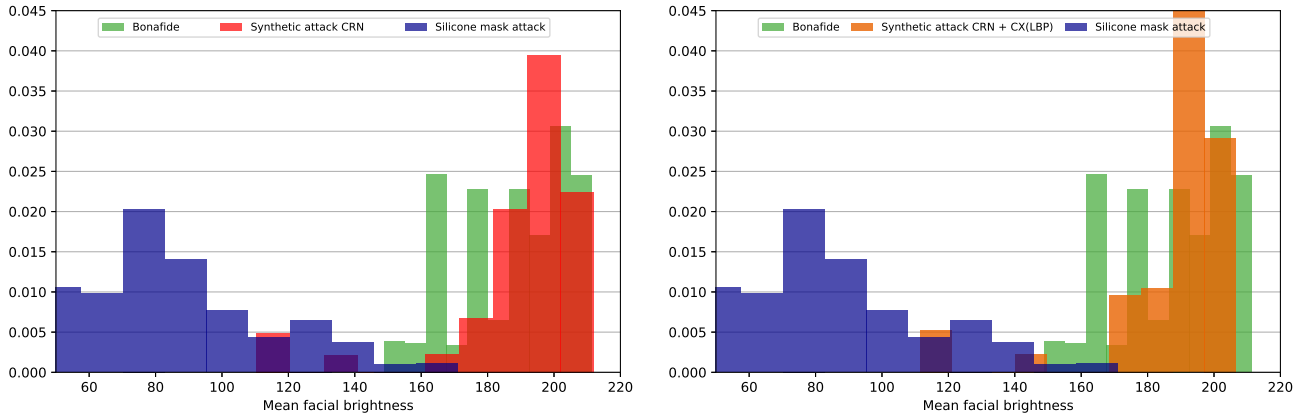
In this section, we carry out a performance evaluation of spoofing attack detection when confronting the new proposed synthetic attack in order to quantify the threat caused by the synthetic attack. First, we present the spoofing attack detection algorithms used for the evaluation. Then, we introduce our experimental setup followed by the reported results and discussion.

5.1. Spoofing attack detection baselines

The selected baselines of spoofing attack detection were previously presented in studies of the robustness of thermal spectrum against spoofing attacks [4, 5, 3, 11].

Mean facial intensity (MFI) As defended in [4, 5], mean facial intensity is a simple but efficient solution to prove the user's liveness in the thermal spectrum. This argument can be endorsed by the fact that face regions are rather bright in the thermal spectrum, while presentation attacks are quite dark since they are at a significantly lower temperature than faces. This is also valid for silicone mask attacks since it is expected that the attack region will be relatively darker than the face region even when worn by the attacker. Mean facial intensity can be used merely as a spoofing attack detection score.

Local Binary Patterns and Logistic Regression (LBP+LR) Local binary patterns (LBP) are used to represent the texture variation between bonafide samples and attack samples. Subsequently, logistic regression (LR)



(a) (b)

Figure 5: Score distribution of mean facial intensity baseline for bonafide and attack samples. **(a)** synthetic attack CRN **(b)** synthetic attack CRN+CX(LBP)

	PSNR	SSIM	Brightness	Contrast
CRN	15.576 (±4.246)	0.610 (±0.103)	0.943 (±0.061)	0.738 (±0.183)
CRN+CX(LBP)	15.616 (±4.208)	0.618 (±0.107)	0.945 (±0.062)	0.747 (±0.210)

Table 1: Quality assessment of the synthetic attacks.

is used to build a classifier to label samples as *bonafide* or *attack*. LBP features are normalised before training the LR model. We have applied normalisation to zero mean and unit standard deviation using parameters extracted only from the bonafide feature set. Given an LR-trained model, the output of this spoofing attack detection is a probability of a sample being bonafide.

5.2. Experiments and results

The performance evaluation of the presented spoofing attack detection baselines is performed on the CSMAD dataset along with the synthetic attacks obtained using the different visible-to-thermal conversion models. CSMAD dataset provides video samples that are split into frames. Spoofing attack detection scores are computed at the frame level.

Face regions are cropped by extracting the face coordinates on the visible spectrum and projecting them on thermal face images. Mean facial intensity is computed across the face region. Figure 5 illustrates the score distribution of mean facial intensity for bonafide samples and attack samples. The score distribution of the bonafide samples is illustrated in green and the silicone mask attack in blue in Figures 5a and 5b. We observe that the two score distributions hardly overlap. However, the score distribution for the synthetic attack generated by the two different models of visible-to-thermal spectrum conversion significantly overlaps with the score distribution of the bonafide samples. Accordingly, we can deduct that the proposed synthetic attack has led to a failure of the spoofing attack detection solution based on mean facial intensity.

For the LBP+LR baseline, we split the CSMAD dataset into 14 partitions, each partition corresponds to a specific subject. For each cross-validation fold, 13 partitions are selected to train the spoofing attack detection model and the remaining partition is used for testing. The splitting of the dataset is defined in a way to ensure a disjoint set of subjects so that the spoofing attack detection model does not learn subject-specific information. Figure 6 presents the detection error tradeoff (DET) curves corresponding to each of the studied attacks. For the silicone mask attack, we observe that the LBP+LR based spoofing attack detection report a considerably low error, reflecting this solution's robustness against silicone mask attacks. The performance of the LBP+LR baseline drastically decreases when dealing with the proposed synthetic attacks. In a scenario of an extremely secure spoofing attack detection system where almost no attacker will be able to breach the system, if we permit a false acceptation rate of 0.1% for instance, we will obtain a false alarm rate of 30-33%. Comparing the performance of the spoofing attack detection solution for the synthetic attack obtained by the two different models of spectrum conversion, we note that combining the CRN module with the loss computed at the LBP map level led to more

	MFI	LBP-LR
Silicone mask attack	2.286	0.207
Synthetic attack CRN	58.543	7.432
Synthetic attack CRN + CX(LBP)	56.513	11.603

Table 2: Equal error rate (%) of spoofing attack detection baselines when exposed to different attacks.

challenging attacks.

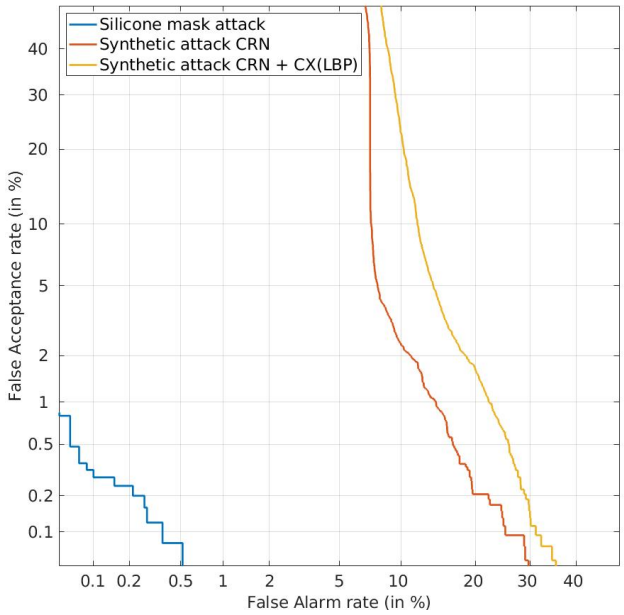


Figure 6: Detection error tradeoff (DET) curves of LBP+LR spoofing attack detection baseline for different attacks.

The equal error rates reported by the two spoofing attack baselines for the different attacks are gathered in Table 2. It is observable that the proposed synthetic attacks represents a considerably high threat, in comparison to silicone mask attack that is considered so far a challenging attack for thermal spectrum. The equal error rate has increased from 2.28% to 58.54% and from 0.20% to 11.60% for mean facial intensity and LBP+LR spoofing attack detection, respectively.

When the attacker does not have any prior knowledge about the spoofing countermeasure implemented in the system, the performance of the spoofing attack detection significantly drops when it faces the synthetic attack generated by CRN model. We perceive that the equal error rate increased from 0.20% to 7.43%. Although when the attacker does indeed have a prior information about the spoofing countermeasure that is being employed, he/she can use this information in a way to customise his/her attack to have higher chances to breach the system. This scenario is executed for visible-to-thermal spectrum conversion model *CRN + CX(LBP)*, where we have used the LBP map information to better attack the LBP+LR based spoofing attack detection system.

It is worth noting that the equal error rate has increased from 0.20% for silicone mask attacks to 11.60% for synthetic thermal attacks using a visible-to-thermal spectrum conversion model that has been trained on a limited number of samples from VIS-TH dataset [16]. Interest in thermal imagery for face biometrics is increasing steadily, thus we can presume that the amount of thermal face data that will be publicly available will grow significantly, which will lead to synthesising highly realistic thermal images which will be difficult to differentiate from bonafide thermal images.

6. Conclusion

Deploying thermal technology in face biometric systems requires an extensive study of its implications and the risk it may encounter. In this paper, we proposed a new attack on thermal face biometric systems, that takes place at the post-sensor level. This thermal attack is generated through visible-to-thermal spectrum conversion of visible face images that are available on the social networks or acquired sneakily from a distance. A quality assessment of the synthetic attacks has been performed by comparing the generated thermal images to thermal bonafide samples. Subsequently, the threat of the proposed synthetic attack was measured through an assessment of the vulnerability of two existing spoofing attack detection solutions, designed for thermal spectrum, to these attacks. This evaluation reported a significant drop in performance of the two used baselines when they face the proposed synthetic attack compared to when they are exposed to silicone mask attack, the most challenging attack for thermal spectrum studied so far. A scenario representing an attacker that has prior knowledge of the spoofing attack detection solution deployed in the biometric system is also explored. For LBP-based spoofing attack detection system, we have adjusted the visible-to-thermal spectrum conversion model in a way that it aims to generate thermal images of which the LBP map is closer to the LBP map of thermal bonafide samples. The obtained synthetic attacks using the customised spectrum conversion model have increased the error rate reported by the targeted spoofing attack detection approach.

We have proven through this study that, even though it is true that thermal spectrum is extremely robust against presentation attacks, this does not deny the fact that new attacks customised for thermal imagery might act as a serious threat.

References

[1] Flir systems. http://www.flir.com/.

[2] ISO/IEC 30107. Information technology biometric presentation attack detection part 1: Framework. *International Organization for Standardization*, Jan 2016.

[3] Akshay Agarwal, Daksha Yadav, Naman Kohli, Richa Singh, Mayank Vatsa, and Afzel Noore. Face presentation attack with latex masks in multispectral videos. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 81–89, 2017.

[4] Sushil Bhattacharjee and Sébastien Marcel. What you can't see can help you-extended-range imaging for 3d-mask presentation attack detection. In *2017 International Conference of the Biometrics Special Interest Group (BIOSIG)*, pages 1–7. IEEE, 2017.

[5] Sushil Bhattacharjee, Amir Mohammadi, and Sébastien Marcel. Spoofing deep face recognition with custom silicone masks. In *2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, pages 1–7. IEEE, 2018.

[6] Qifeng Chen and Vladlen Koltun. Photographic image synthesis with cascaded refinement networks. In *IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy, October 22-29*, pages 1520–1529. IEEE Computer Society, 2017.

[7] Valeria Chiesa and Jean-Luc Dugelay. Advanced face presentation attack detection on light field database. In *2018 International Conference of the Biometrics Special Interest Group (BIOSIG)*, pages 1–4. IEEE, 2018.

[8] Tejas I Dhamecha, Aastha Nigam, Richa Singh, and Mayank Vatsa. Disguise detection and face recognition in visible and thermal spectrums. In *2013 International Conference on Biometrics (ICB)*, pages 1–8. IEEE, 2013.

[9] Nesli Erdogmus, Neslihan Kose, and Jean-Luc Dugelay. Impact analysis of nose alterations on 2d and 3d face recognition. In *2012 IEEE 14th International Workshop on Multimedia Signal Processing (MMSP)*, pages 354–359. IEEE, 2012.

[10] Leon A. Gatys, Alexander S. Ecker, and Matthias Bethge. Image style transfer using convolutional neural networks. In *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*, pages 2414–2423. IEEE Computer Society, 2016.

[11] Anjith George, Zohreh Mostaani, David Geissenbuhler, Olegs Nikisins, André Anjos, and Sébastien Marcel. Biometric face presentation attack detection with multi-channel convolutional neural network. *IEEE Transactions on Information Forensics and Security*, 15:42–55, 2019.

[12] Anil K Jain, Karthik Nandakumar, and Abhishek Nagar. Biometric template security. *EURASIP Journal on advances in signal processing*, 2008:1–17, 2008.

[13] Sooyeon Kim, Yuseok Ban, and Sangyoun Lee. Face liveness detection using a light field camera. *Sensors*, 14(12):22471–22499, 2014.

[14] Khawla Mallat, Naser Damer, Fadi Boutros, and Jean-Luc Dugelay. Robust face authentication based on dynamic quality-weighted comparison of visible and thermal-to-visible images to visible enrollments. In *2019 22th International Conference on Information Fusion (FUSION)*, pages 1–8. IEEE, 2019.

[15] Khawla Mallat, Naser Damer, Fadi Boutros, Arjan Kuijper, and Jean-Luc Dugelay. Cross-spectrum thermal to visible face recognition based on cascaded image synthesis. In *2019 International Conference on Biometrics (ICB)*, pages 1–8. IEEE, 2019.

[16] Khawla Mallat and Jean-Luc Dugelay. A benchmark database of visible and thermal paired face images across multiple variations. In *2018 International Conference of the Biometrics Special Interest Group (BIOSIG)*, pages 1–5. IEEE, 2018.

[17] Khawla Mallat and Jean-Luc Dugelay. Facial landmark detection on thermal data via fully annotated visible-to-thermal data synthesis. In *submitted to 2020 International Joint Conference on Biometrics (IJCB)*, 2020.

[18] Sébastien Marcel, Mark S Nixon, and Stan Z Li. *Handbook of biometric anti-spoofing*, volume 1. Springer, 2014.

[19] Roey Mechrez, Itamar Talmi, and Lihi Zelnik-Manor. The contextual loss for image transformation with non-aligned data. In *Computer Vision - ECCV 2018 - 15th European Conference, Munich, Germany, September 8-14, 2018, Proceedings*, volume 11218 of *Lecture Notes in Computer Science*, pages 800–815. Springer, 2018.

[20] Michael O Mireku, Tessa R Flack, Kay L Ritchie, et al. Face morphing attacks: Investigating detection with humans and computers. *Cognitive Research*, 4(1), 2019.

[21] Amir Mohammadi, Sushil Bhattacharjee, and Sébastien Marcel. Deeply vulnerable: a study of the robustness of face recognition to presentation attacks. *IET Biometrics*, 7(1):15–26, 2017.

[22] Timo Ojala, Matti Pietikäinen, and David Harwood. A comparative study of texture measures with classification based on featured distributions. *Pattern recognition*, 29(1):51–59, 1996.

[23] Ioannis Pavlidis and Peter Symosek. The imaging issue in an automatic face/disguise detection system. In *Proceedings IEEE Workshop on Computer Vision Beyond the Visible Spectrum: Methods and Applications (Cat. No. PR00640)*, pages 15–24. IEEE, 2000.

[24] Nalini K Ratha, Jonathan H Connell, and Ruud M Bolle. An analysis of minutiae matching strength. In *International Conference on Audio-and Video-Based Biometric Person Authentication*, pages 223–228. Springer, 2001.

[25] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *CoRR*, abs/1409.1556, 2014.

[26] Fatemeh Vakhshiteh, Raghavendra Ramachandra, and Ahmad Nickabadi. Threat of adversarial attacks on face recognition: A comprehensive survey. *arXiv preprint arXiv:2007.11709*, 2020.