Efficient integration of thermal technology in facial image processing through interspectral synthesis

Dissertation

submitted to

Sorbonne Université

in partial fulfilment of the requirements for the degree of
Doctor of Philosophy

Author:
Khawla MALLAT

Scheduled for defense on
7th of July, 2020
before a committee composed of:

Thesis advisor	**Prof. Jean-Luc DUGELAY**, EURECOM, France
Reviewers	**Prof. Boulbaba BEN AMOR**, IIAI, UAE
	Prof. Hazım Kemal EKENEL, Istanbul Technical University, Turkey
Examiners	**Prof. Bernard MERIALDO**, EURECOM, FRANCE
	Prof. Bernadette DORIZZI, Télécom SudParis, France
	Dr. Cécile ICHARD, GTD internationals, France

Abstract

Face biometric systems are now a reality in numerous mainstream applications including access control, banking and forensics. Notably, face recognition systems have recently advanced and achieved striking performances due to the uprise of deep learning and the the abundant, almost endless, amount of available training data. However, these systems, that are mainly deployed in visible spectrum, are subject to fail when employed in unconstrained scenarios. Among the main challenges in visible spectrum based systems, variable or low illumination conditions have been proved to be some of their major weaknesses. A promising approach to acquire crisp images in total darkness is using thermal imagery. Thermal imaging technology has significantly evolved during the last couple of decades, mostly thanks to thermal cameras having become more affordable and user friendly. However, and given that the exploration of thermal imagery is reasonably new, only a few public databases are available to the research community. This limitation consequently prevents the impact of deep learning technologies from generating improved and reliable face recognition systems that operate in the thermal spectrum. A possible solution relates to the development of technologies that bridge the gap between visible and thermal spectra. In attempting to respond to this necessity, the research presented in this dissertation aims to explore interspectral synthesis as a direction for efficient and prompt integration of thermal technology in already deployed face biometric systems.

As a first contribution, a new database, containing paired visible and thermal face images, which was acquired with a dual camera that allowed for the simultaneous capture of face images in both spectra, was collected and made publicly available to foster research in thermal face image processing. Motivated by the need for fast and straightforward integration into existing face recognition systems, a following contribution consisted in proposing a cross-spectrum face recognition framework based on a novel approach of thermal-to-visible face synthesis in order to estimate the visible face from the thermal input, when the visible image cannot be provided, e.g. in poorly lit environments. The proposed approach is based on deep generative models and was trained on a set of paired visible and thermal data to learn a mapping from the thermal face to its visible equivalent. After this initial work, another contribution presents the development of an illumination invariant face recognition system that incorporates a novel, dynamic

quality-weighted, fusion of visible and thermal spectrum at the score level. Thanks to the proposed mechanism, an uninterrupted and efficient functioning of a face recognition setup during day and night time may be ensured.

Motivated by the favorable results achieved in the first part of our research work, additional contributions presented in this thesis explore the process of interspectral synthesis in the reverse direction, i.e. from visible to thermal spectrum. Visible-to-thermal image synthesis was employed to address the shortage of annotated public face databases in thermal spectrum, which limits the development of fundamental task in thermal face image processing. With the scope of this study being focused on the facial landmark detection task, fully annotated synthesized thermal face databases were obtained by transforming public annotated visible face databases into thermal spectrum. Facial landmark detectors trained on the synthesized thermal face databases led to significant improvements in landmark detection accuracy. A final contribution explored visible-to-thermal synthesis to study the impact of spoofing attacks on thermal face biometric systems. The robustness of thermal based systems lies in the acquisition process itself as it provides proof of liveness by detecting the heat emitted by the face. A new thermal attack, at the post-sensor level, is then proposed. Thermal face images, that are obtained by visible-to-thermal face synthesis, are directly injected in the communication channel after the sensor. In order to increase the difficulty of the proposed setup, a scenario where the attacker has a priori knowledge about the spoofing countermeasure employed by the system is also considered. Such a priori knowledge is exploited in order to synthesize more threatening attacks for a given countermeasure technique. The evaluation of spoofing detection systems when facing the proposed attack highlights the vulnerability of thermal face recognition systems to the proposed indirect attack.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Context and motivation

Biometric recognition is rapidly emerging as a reliable and a fast tool of identity management by analyzing physical and behavioral characteristics specific to each individual that are distinctive, permanent and universal. While until very recently fingerprint was known to be the most prevalent form of biometrics in commercial biometric systems [1], face is now taking over to establish itself as a more convenient and accessible alternative. Face represents the most natural and intuitive mean of recognition by humans, and the information conveyed in face is especially rich and diverse. Unlike iris, hand geometry and hand veins biometric systems, face recognition does not require costly and high accuracy acquisition sensors. Furthermore, face recognition does not involve physical interaction with the end user, thereby facilitating the identification of target subjects from relatively great distances without the target's cooperation, a significant asset for law enforcement and security applications.

Over three decades of extensive research has led to a massive deployment of face recognition systems along with substantial gains and improvements in performance. This is due to a variety of factors that include the steady hardware developments and the outbreak of abundant face data at the disposal of researchers. Face recognition systems spans nowadays a wide range of vertical industries including banking, border control, healthcare and security applications. Following the explotion in the ubiquity of smart devices equipped with camera sensors, face recognition is now powering through *Internet of Things* market.

In spite of its world-wide deployment and its growing popularity, face recognition systems are still prone to fail when employed in unconstrained conditions. Face recognition

systems are exclusively deployed using 2D and/or 3D acquisition sensors operating solely in visible spectrum that suffers from various limitations. Among the main challenges confronted by visible face recognition systems, variable or low illumination conditions have been proved to be some of the major weaknesses of such systems [2, 3, 4], due to the reflective nature of visible spectrum. Furthermore, head pose [5], facial expression [6], makeup [7] are only some of other challenges that decreases the reliability of visible face recognition systems. Moreover, visible face recognition systems are also threatened by presentation attacks that endeavor to spoof the system and gain unauthorized access [8, 9, 10, 11]. Some prompt actions have been taken such as requiring an eye blink, smile or other visual reaction to prove the liveness of the user, yet this can be easily tricked using video replay attacks. Presentation attack detection [12, 13, 14] is still a very active research area, although visible face recognition systems are extensively implemented for border and access control and surveillance systems. Thereby, it is necessary to seek solutions that are cost effective and easy to integrate with existing face recognition systems.

Thermal face recognition has emerged as a promising complement to visible face recognition, as it provides efficient solutions to tackle the challenges encountered by systems based on visible spectrum. Thermal face images are invariant to light changes due to the fact that the radiation detected by the thermal sensor is directly emitted by the human face [15], and not reflected as it is the case for visible spectrum. Therefore, it is possible to acquire a crisp thermal image without any external source of illumination, based on subtle differences in temperature. Moreover, the sensitivity of visible face recognition systems to head pose, facial expression and makeup variations is partly due to the change of the reflectance of visible light, this is however not an issue in thermal spectrum. Thereby, thermal face recognition systems are less affected by these variations [16]. Additionally, thermal technology can be used as a presentation attack detection tool, as it provides an evidence of the user's liveness by simple acquisition.

Thermal imagery was initially limited to military use. The first thermal line scanner was developed in 1947 by the US military and took one hour to produce one single image [17]. In 1966, the first real-time commercial thermal imager was launched. By the end of the 1990s, uncooled focal plane arrays with higher resolution were introduced at a reduced prices, which motivated their use in civil applications. These applications include building and roof inspection, environment control, medical testing and diagnosis and art analysis [18]. However, the cost of thermal sensors remained exorbitantly high and the quality of thermal data was insufficient for thermal spectrum to be explored in face recognition applications. During the last decade, driven by the progress in microelectronics and the dramatic lowering of manufacturing prices, uncooled microbolometers focal plane arrays are providing high thermal sensitivity and high spatial resolution at very

competitive prices. This even pertains to some models of smart phones that are starting to be equipped with thermal imagers [19, 20]. Consequently, research interest in thermal face recognition has significantly grown. However, the data in thermal spectrum available for the research community has not increased at a comparable pace to that of visible spectrum face databases. This is a limitation for thermal spectrum investigation as a biometric, particularly in the context of the current deep learning based trends, which tend to be particularly data hungry. While visible face databases are abundant and can lead to the training of highly complex deep neural models, the same cannot be done, as of the time of writing of this dissertation, for thermal imagery.

While it is obvious that the dropping manufacturing costs in thermal sensors will eventually make those capturing devices as mainstream as those in visible spectrum, security related scenarios in which thermal sensors are already a reality cannot wait for the available resources in thermal spectrum to balance with that of the visible spectrum. For thermal spectrum databases to leverage the potential of those deep learning based algorithms, characterisable by their data needy functioning, methods that allow to exploit the complementary of the information that lies in both thermal and visible spectra need to be developed, motivating the research presented in this dissertation.

The principal contributions of the presented work are focused on the development of new advances that lay the ground for an efficient and prompt integration of thermal technology in already commercially deployed face biometric systems. Such contributions are needed to lead a step up in directing the development of state-of-the-art in thermal facial image processing and sustain the growing usage of thermal spectrum. Promoting the integration of thermal technology in existing face biometric systems is based on the use of interspectral synthesis in both directions. Thermal face images can be transformed in visible spectrum, bridging the impact of the aforementioned factors on visible faces, and can then benefit from the wide range of available facial image processing tools. Alternatively, it is possible to generate synthetic thermal face images required for the design and the development of a specific task, by transforming available visible face databases to thermal spectrum. The array of the proposed solutions throughout this thesis prevents the adaptation and re-optimization of available resources to operate on thermal spectrum, as well as the extensive collection of thermal face databases, that can be costly and inefficiently time consuming.

1.2 Contributions and thesis outline

The need for the availability of multi spectral resources while massive thermal data is not at the disposal for the research community has motivated the lines of research that are

presented in this dissertation and outlined in the following paragraphs. The contributions made by the research included in this thesis are then as follows.

Chapter 3

A first contribution is that of Chapter 3, which presents the efforts developed by the author of this dissertation in collecting a dual visible-thermal, paired-by-design face database that include numerous variations in terms of facial expression, head pose, occlusion and illumination conditions to replicate real-life, challenging scenarios for the face recognition state-of-the-art systems. The careful design of this database aims to foster the research in the field in as much as it also provides with the foundations on which the remainder of the works presented in this thesis are built. Besides the design and discussion of the collection protocol for the database, results of initial experiments for its validation in face recognition research were reported.

Part of the work presented in this chapter was published in:

- **K. Mallat**, J.-L Dugelay, "**A benchmark database of visible and thermal paired face images across multiple variations**" in *Proc. 17th International Conference of the Biometrics Special Interest Group BIOSIG*, Darmstadt, Germany, September 2018.

which was awarded with the best poster award. The VIS-TH database has since then been available to the research community and has been downloaded by over 25 teams worldwide.

Chapter 4

The work presented in Chapter 4 relates to the first application of state-of-the-art deep generative models to the problem of thermal-to-visible data synthesis. Recent advancements in deep learning have led to the development of deep neural network topologies capable of generating high-quality transformation between images of a significantly different domains, with our interested being focused on cascaded refinement networks (CRNs) [21]. In particular, our work puts the focus on the application of CRNs to the problem of cross-spectrum face recognition in highly challenging scenarios, i.e. the absolute darkness scenario, by allowing for thermal data to be immediately usable by visible spectrum based systems by means of a CRN-based transformation. This contribution prevents the extensive recollection of enrollment data in thermal spectrum and the development of reliable algorithms for thermal face recognition. Results validated

the proposed methodology and opened the door to the exploration of the better use of a CRN-based transformation to further face image processing related tasks in the remainder of this thesis.

Part of the work presented in this chapter was published in:

- **K. Mallat**, N. Damer, F. Boutros, A. Kuijper, J.-L Dugelay, "**Cross-spectrum thermal to visible face recognition based on cascaded image synthesis**" in *Proc. best conference*, in *Proc. 12th IARP International Conference on Biometrics ICB*,Crete, Greece, June 2019.

Chapter 5

Motivated by the positive results in Chapter 4, Chapter 5 reports the investigation of mechanisms that intelligently incorporate the best attributes of face recognition systems that work simultaneously on (i) visible images, and (ii)on synthesized thermal-to-visible images. Whilst the quality achieved by thermal-to-visible face synthesis via the method reported in Chapter 4 achieves a high quality and realism, of particular benefit in scenarios in which the visible spectrum cannot cope, i.e. in poorly lit environments, the resulting images are evidently a few steps behind that of standard, visible spectrum, face images. The main contribution of this chapter then relates to the development of a novel method based on dynamic fusion of matching scores of visible probes and synthesized thermal-to-visible probes against visible gallery, via the usage of various quality metrics widely used in image processing. The proposed method allows for a face recognition system to smoothly transition between using visible or synthesized thermal-to-visible images depending the relevance of each sample determined by a quality score. The presented contribution enabled the design of illumination invariant face recognition system, by exploiting the invariance of thermal spectrum to illumination changes, without the requirement of thermal specific face recognition algorithms.

Part of the work presented in this chapter was published in:

- **K. Mallat**, N. Damer, F. Boutros, J.-L Dugelay, "**Robust face authentication based on dynamic quality-weighted comparison of visible and thermal-to-visible images to visible enrollments**" in *Proc. 22nd International Conference on Information Fusion FUSION*,Ottawa, Canada, July 2019.

Chapter 6

Following, Chapter 6 explores the potential benefit of CRN-based interspectral synthesis for a task related to, but different than face recognition, that is of facial landmark detection. The contribution of this work consists in introducing an innovative concept, that to the our knowledge, hasn't been previously explored in the literature. This novel concept aims to tackle the shortage of annotated data in thermal spectrum. Given the positive results achieved in the experiments included in the other chapters of this thesis, we propose the leveraging of CRN-based image synthesis in the reverse spectral direction, i.e. from visible to thermal spectrum, in order to synthesize thermal face databases and exploit the annotation information provided in the visible spectrum for facial landmark detection. The presumably higher information domain of visible spectrum compared to that of thermal domain allows for the resulting transformation to be of extremely high quality. Relating to our new application of interest of facial landmark detection, the resulting high-quality, synthesized, thermal face databases allow for the training of facial landmark detectors directly on the thermal spectrum. Facial landmark detection in thermal spectrum still remains a challenge, mainly due to the limited resources of databases with annotated landmarks in the thermal spectrum. Our proposed approach achieves remarkable results with high facial landmark detection accuracy evaluated on thermal data of different quality.

Part of the work presented in this chapter was submitted to:

- **K. Mallat**, J.-L Dugelay "**Facial landmark detection on thermal data via fully annotated visible-to-thermal data synthesis**" in *Submitted to International Joint Conference on Biometrics IJCB*, Houston, USA, September 2020.

Chapter 7

The last contribution, presented in Chapter 7 of this dissertation, relates to a consequence of the great success that face image processing techniques are acquiring in the recent years. Face recognition systems are now widely used by both public authorities and domestic users. Consequently, and whilst these methods normally provide with an enhanced level of security in the authentication process for the average user, spoofing attacks have become increasingly common, attracting wide research interest for face recognition among many other biometric traits. Thermal imagery is generally considered as a natural spoofing countermeasure. However, its robustness to spoofing threats lies in the acquisition process itself. In the work presented in Chapter 7, we take the role of an attacker that intends to break a thermal face biometric system by short-circuiting the thermal sensor and injecting a thermal face image in the communication channel between the sensor and

the subsequent processing module. The proposed attack, performed at post-sensor level, is obtained by visible-to-thermal face synthesis. Two spoofing scenarios are studied: (i) the attacker blindly injects a synthesized thermal image, or (ii) the attacker possesses a prior knowledge about the spoofing countermeasure of the target system. For the second scenario, a customized interspectral synthesis model, that incorporates the prior information in the development of visible-to-thermal face synthesis, is introduced in order to leverage more robust attacks against the targeted spoofing countermeasures. While initial results in the literature report for thermal imagery to be a very robust countermeasure to presentation attacks, the work presented in this dissertation highlights the vulnerability of spoofing countermeasures when confronting attacks at post-sensor level. This contribution aims to study the vulnerability of thermal face biometric systems and the threats it may potentially confront once it is deployed.

Part of the work presented in this chapter was submitted to:

- **K. Mallat**, J.-L Dugelay, "**Indirect synthetic attack on thermal face biometric systems via visible-to-thermal spectrum conversion**" in *Submitted to 25th International Conference on Pattern Recognition ICPR*, Milan, Italy, January 2021.

The overall outline of this thesis is appended by means of a literature review relating to facial image processing and thermal imagery in **Chapter 2**, and conclusions and future work, which are presented in **Chapter 8**.

Publications

Discussed in this manuscript

1. **K. Mallat**, J.-L Dugelay, "**Indirect synthetic attack on thermal face biometric systems via visible-to-thermal spectrum conversion**" in *Submitted to 25th International Conference on Pattern Recognition ICPR*, Milan, Italy, January 2021. (under review)

2. **K. Mallat**, J.-L Dugelay "**Facial landmark detection on thermal data via fully annotated visible-to-thermal data synthesis**" in *Submitted to International Joint Conference on Biometrics IJCB*, Houston, USA, September 2020. (under review)

3. **K. Mallat**, J.-L Dugelay, "**Conversion thermique-visible en imagerie faciale**" in *Proc. Compression et REprésentation des Signaux Audiovisuels CORESA*, Sophia Antipolis, France. 2020.

4. **K. Mallat**, N. Damer, F. Boutros, J.-L Dugelay, "**Robust face authentication based on dynamic quality-weighted comparison of visible and thermal-to-visible images to visible enrollments**" in *Proc. 22nd International Conference on Information Fusion FUSION*, Ottawa, Canada, July 2019.

5. **K. Mallat**, N. Damer, F. Boutros, A. Kuijper, J.-L Dugelay, "**Cross-spectrum thermal to visible face recognition based on cascaded image synthesis**" in *Proc. best conference*, in *Proc. 12th IARP International Conference on Biometrics ICB*, Crete, Greece, June 2019.

6. **K. Mallat**, J.-L Dugelay, "**A benchmark database of visible and thermal paired face images across multiple variations**" in *Proc. 17th International Conference of the Biometrics Special Interest Group BIOSIG*, Darmstadt, Germany, September 2018.

Other works

7. N. Damer, F. Boutros, **K. Mallat**, F. Kirchbuchner, J.-L. Dugelay, A. Kuijper, "**Cascaded Generation of High-quality Color Visible Face Images from Thermal Captures**" in *arXiv preprint arXiv:1910.09524* submitted on Arxiv October 2019.

8. **K. Mallat**, S. K. Datta, J.-L Dugelay, "**IoT Based People Detection for Emergency Scenarios**" in *IEEE 9th International Conference on Consumer Electronics (ICCE-Berlin)*, Berlin, Germany, September 2019.

9. **K. Mallat**, C. Galdi, J.-L Dugelay, "**Evaluation of the facial makeup impact on femininity appearance based on automatic prediction**" in *2nd Cosmetic Measurement & Testing Symposium*, Cergy-Pontoise, France, June 2016.

Awards

- **3MT contest runner-up Award (2nd prize)**, for the work entitled **"Spectrum conversion from thermal to visible images: Safety and security applications"** at the *27th European Signal Processing Conference, EUSIPCO*. A Coruña, Spain. September 2019.

- **Best poster award** for the poster entitled **"A benchmark database of visible and thermal paired face images across multiple variations"** at the *17th edition of the International Conference of the Biometrics Special Interest Group (BIOSIG)*. Darmstadt, Germany. September 2018.

Chapter 2

Thermal spectrum in facial image processing: literature overview

This chapter reveals the motivation behind the usage of thermal spectrum in facial image spectrum. An overview of the literature of relevance to facial image processing in thermal spectrum is provided. This includes a review of research works that study thermal facial image processing under unconstrained scenarios. Further detailed read can be found in widely cited survey articles for thermal spectrum in face biometric systems [16, 22].

2.1 Spectral imaging

Spectral imaging refers to imaging methods which operate in different bands of the electromagnetic spectrum. In this section, some background fundamentals related to the electromagnetic spectrum and infrared band are presented. The motivation behind the usage of infrared spectrum, in particular thermal spectrum, in facial image processing is then defended.

2.1.1 Electromagnetic spectrum

Electromagnetic radiation is a form of energy that propagates through space as electromagnetic waves carrying packets of energy called photons or light quanta [23]. The electromagnetic energy spans a broad range of wavelengths and frequencies, known as

the electromagnetic spectrum. The EM spectrum is usually divided into separate bands, illustrated in Figure 2.1, based on different characteristics of emission, transmission and absorption of each band.

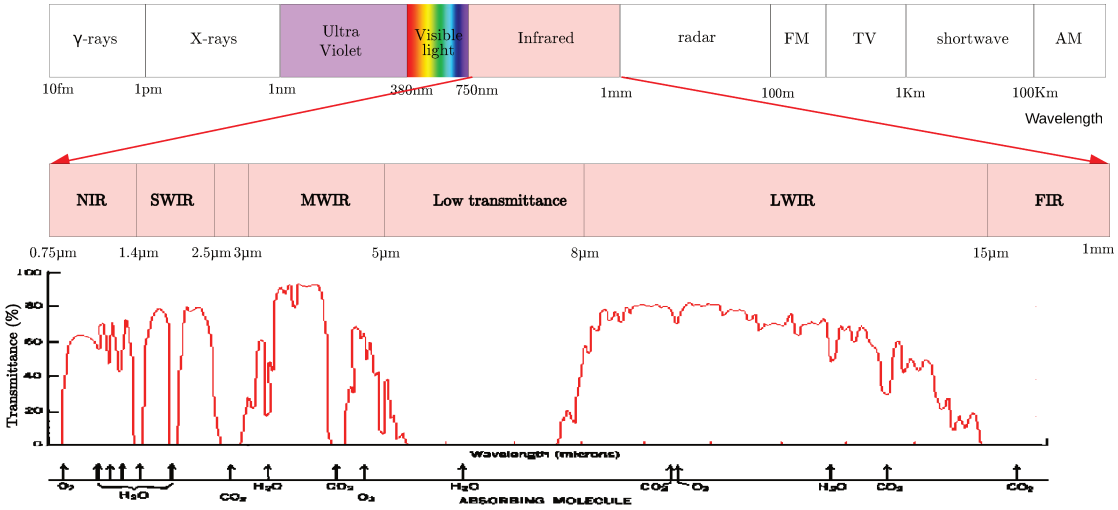


Figure 2.1: Electromagnetic spectrum: bands and their corresponding wavelengths. Spotlight on the infrared band and the corresponding atmospheric transmittance window. Figure adapted from [24].

The visible spectrum is the band of the EM spectrum that is visible for the human eye. The visible band corresponds to a narrow range of wavelengths spanning from 380nm to 740nm. Each wavelength of the visible light band matches a particular color. Objects do not in fact have colors, yet they have properties that indicate which wavelengths are absorbed and which are reflected. The human vision system, much like visible sensors, are sensitive to the reflected light wavelengths of the scene to construct an image. While most mammals are also sensitive to visible light, many other species have the ability to see outside the visible spectrum. Some insects can see in ultraviolet spectrum, which enables them to detect nectar in flowers. Also, birds can see in the ultraviolet spectrum. In fact, they have gender-dependent patterns marked on their feathers that are only perceptible in ultraviolet light. Other species such as mosquitoes, bats, and some snakes, however, can use sub-bands of the infrared spectrum for vision.

2.1.2 Infrared spectrum

Infrared imagery has been widely used mainly due to the advantages it offers over visible imagery, notably for facial image processing. Face images in infrared spectrum can be acquired in any illumination condition. In addition, subcutaneous information of faces can be extracted using infrared spectrum. Infrared spectrum is also less sensitive to scattering and absorption by smoke, dust or fog compared to reflected visible light.

According to the International Commission on Illumination [25], it is recommended to divide the infrared spectrum into four sub-bands as shown in table 2.1:

IR sub-bands	Acronym	Wavelength
near IR	NIR	0.75 - 1.4
short wave IR	SWIR	1.4 - 3
medium wave IR	MWIR	3 - 5
long wave IR	LWIR	8 - 15

Table 2.1: Spectral decomposition of infrared spectrum according to International Commission on Illumination [25].

Each sub-band corresponds to continuous frequency block of the solar spectrum which are divided by absorption lines of different atmospheric gazes [26], as depicted in figure 2.1. Most of the infrared spectrum is not usable as it is blocked by the atmosphere. Also, a window of the infrared spectrum between MWIR and LWIR from 5µm to 8µm has no atmospheric transmission.

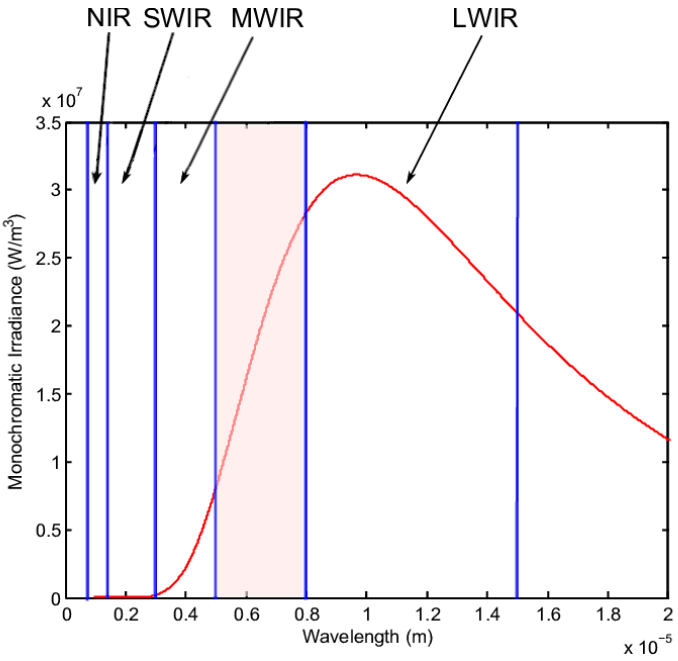


Figure 2.2: Heat emission by the human body predicted by Planck's law at 305 K [16]. The highlighted part represents the dead zone with no atmospheric infrared transmission.

According to the Planck's law, each body being in the thermal equilibrium emits radiation in a broad spectral range. In the context of face processing, the difference between different infrared sub-bands originates as a consequence of the human body's heat emission, as represented in Figure 2.2. The most of the heat energy is emitted in

the LWIR range, therefore it is referred to as the thermal sub-band. To a lesser degree, significant amount of heat is also emitted in MWIR sub-band, for this reason the term 'thermal spectrum' can sometimes be extended to include MWIR sub-band. LWIR and MWIR sub-bands can be used to passively detect thermal emissions of the face without requiring an external source of illumination. Whereas NIR and SWIR require appropriate illumination as the facial heat emission is nearly inexistent in these sub-bands. Figure 2.3 displays face images in visible spectrum and in different sub-bands of infrared spectrum. NIR and SWIR face images seem more similar to the visible spectrum image than MWIR and LWIR images. This is due to the fact that visible, NIR and SWIR spectra acquire reflected radiation, whereas MWIR and LWIR acquire emitted radiations of the face.

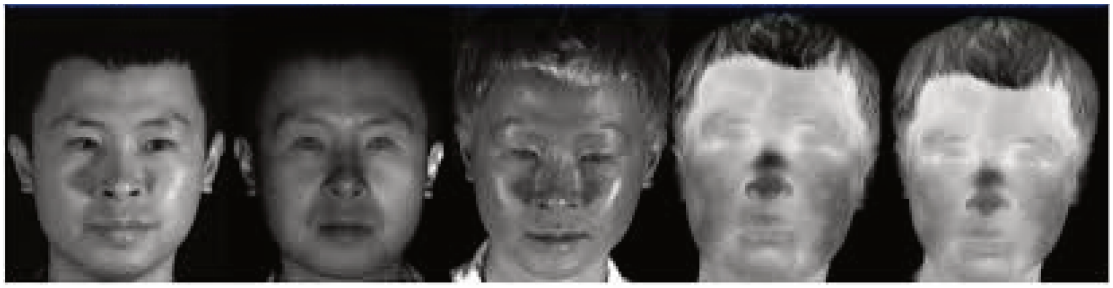


Figure 2.3: Face images acquired in different spectra. Figure reproduced from [27].

2.2 Thermal spectrum for facial image processing

Thermal image of the human face presents a unique thermal signature which can be used for facial recognition [28], as the facial temperature distribution exhibits individual patterns. Thermal imagery received a lot of attention in face recognition mainly due to the fact that it relies on passive heat radiation and does not need illumination source. The acquisition of a scene depends on the specifications of the thermal sensor, the emissivity of the different objects present in the scene and the temperature difference between them. Noise equivalent temperature difference (NETD) identifies the minimum temperature difference that is required for an object to be separated from the noise, it means that objects with temperature difference below the NETD value will disappear in the noise [29]. NETD is considered as the most common measure to characterize the performance of thermal sensors. Lower values of NETD imply higher sensor performance. Figure 2.4 displays face images acquired with sensors of different NETD values. One can observe that the lower the NETD value is, the higher the quality of the image. However, NETD can be measured using different techniques and under different conditions, which makes it difficult to compare the performance of thermal sensors directly based on the NETD values. Thermal sensor equipped with uncooled micro-bolometer focal plane

arrays are generally characterized by NETD values between 30 and 130 mK (millikelvin), whilst sensors equipped with a cooler may have an NETD value below 20 mK. Although, cooling devices are extremely expensive.

Figure 2.4: Face images acquired with sensors of different values of NETD.

The usage of thermal spectrum in facial image processing grants some advantages over the visible spectrum that can overcome some of the main constraints encountered by visible face recognition systems. However, thermal spectrum suffers from various challenges that originate from the fact that the heat emitted by the human face is affected by a number of factors. The advantages and the limitations of the usage of thermal face images under real-world challenges are discussed in this section. A literature overview of the research studies associated with each challenge is presented.

2.2.1 Illumination variation

As previously stated, thermal sensors acquire face images in a passive manner through sensing the facial thermal emission without the need for external source of illumination. This property of thermal spectrum is the main reason why it has been receiving a lot of attention in facial image processing. The immunity of thermal spectrum to illumination variations tackles the major constrain confronted in visible face recognition. Figure 2.5 shows the impact of illumination variation on visible and thermal spectra. We can see that thermal faces can still be acquired even in total darkness.

Several studies [28,30,31,32,33,34,35,36,37] have proposed the use thermal spectrum to overcome the illumination challenge. Socolinsky et al. [30] introduced a decision based fusion using a weighted combination of visible and thermal matching scores. The introduced approach was evaluated in indoors and outdoors scenarios and proved efficiency in most of the use cases but it failed under extreme illumination conditions. Bhowmik et al. [31] proposed an optimum level of pixel fusion from visible and thermal face images by fusing images as a weighted sum and then projected into eigenspace. The fused eigenfaces are classified using train Multilayer Perceptron. Arandjelovic et al [32,33,34]

Figure 2.5: Impact of illumination variation on visible and on thermal spectrum.

debated that the optimal weights in decision level fusion are illumination related. The authors proposed to fuse matching scores of raw appearance and filtered appearance of the visible and the thermal spectra. The proposed approach is based on the observation that if the best matching is achieved in visible spectrum it is because the illumination change between the gallery and the probe sample is minor and more weight should be associated to the visible spectrum and vice versa. A similar approach was introduced by Moon et al [35] where the fusion of visible and thermal spectra is performed through representing face images by the coefficients obtained from a wavelet decomposition. Other studies [36, 37] using wavelet based fusion schemes were also proposed.

2.2.2 Expression variation

While facial expression still remains a significant challenge for face recognition in visible spectrum, thermal spectrum seems to be less affected by facial expression changes. Due to the reflective nature of visible spectrum, facial expression change yield to a change in light distribution across the face resulting from varying surface normals. This does not impact the thermal spectrum as it detects the heat emitted by the face. Socolinsky et al. [4] have carried out a comparative study of different face recognition approaches in visible and thermal spectra. Experiments performed under facial expression variation showed that face recognition performance on the visible spectrum is always inferior to the performance on thermal spectrum. Kong et al. [38] conducted an extensive study of multi-scale fusion of visible and thermal spectra which showed that that thermal face recognition performed better than visible face recognition under various facial expression conditions. Hariharan et al. [39] introduced a data level fusion scheme that generates an image that contains information from both visible and thermal spectra. The approach is based on empirical mode decomposition. Face recognition experiments proved that thermal spectrum is not affected by facial expression variation as much as the visible

spectrum. The invariance of thermal spectrum to facial expression changes is the reason why emotion recognition is not being widely investigated in thermal spectrum.

2.2.3 Head pose variation

Changes in head pose yield to a change of light distribution across the face and the appearance of shadows that occlude facial features in visible spectrum. Being invariant to light changes, thermal spectrum is less affected by head pose variation compared to visible spectrum. Friedrich et al. [28] proved this by comparing image space differences of thermal and visible spectrum. Abidi et al. [40] studied the fusion of visible and thermal spectrum at data level and at decision level as a solution for a robust face recognition against pose variation by exploiting the thermal information. Pop et al. [41] proposed a score based fusion of visible and thermal spectrum using PCA feature extraction and nearest neighbor classification. The proposed approach improves the face recognition performance reported on visible spectrum.

Being less affected by head pose changes than visible spectrum, several studies [42, 43, 44, 45] have focused solely on thermal spectrum to develop solutions of pose invariant face recognition. Zaeri et al. [42] introduced a new approach for thermal face recognition based on affine moment invariants technique. Face images are divided into 16 non-overlapped components. Similarity measures of the feature vectors corresponding to the different components are fused to obtain a final score. Experimental results have showed that this technique has delivered robustness against head pose variation. Buddharaju et al. [45] proposed using the physiological properties of the human face captured in thermal spectrum. The proposed approach is based on extracting of the vascular network of the face. To generalize the approach to different head pose variation, the vascular network was extracted from images of faces in 5 different poses. The branching points of the skeletonized vascular network are then matched to report face matching scores.

2.2.4 Eyeglasses challenge

Eyeglasses are opaque to the infrared spectrum in the SWIR, MWIR and LWIR sub-bands [16], as the eyeglasses block the emitted radiation. Contrarily, the impact of eyeglasses on the appearance in the visible spectrum is way less significant. Figure 2.6 illustrates the impact of eyeglasses on visible and thermal spectrum.

A lot of efforts were devoted to tackle the eyeglasses challenge in thermal face recognition. Studies conducted by Gyaourova et al [46] and Singh et al. [47] proposed a data level fusion technique based on feature selection in visible and thermal thermal

Figure 2.6: Impact of eyeglasses on visible and on thermal spectrum.

spectrum using a genetic algorithm. Using Haar wavelet and eigen component based features, the proposed fusion technique yielded to a higher performance compared to pure visible or pure thermal face matching, specifically when subjects are wearing eyeglasses. Heo et al. [48] studied fusion techniques at data level and at decision level, while proposing to replace the detected eyeglasses by a generic eye template. The eyeglasses replacement resulted in significant improvement in face recognition performance A similar solution was proposed by Kong et al. [38] where the eyeglasses are detected and replaced by the average eye appearance. Wong et al. [49] used the face reconstruction information from the visible image to replace the eyeglasses patches in thermal spectrum.

2.2.5 Presentation attacks

One of the main advantages of thermal spectrum over visible spectrum in facial biometric systems is its robustness to presentation attacks. Presentation attack consists in presenting a fake human biometric sample in attempt to gain unauthorized access or evade biometric recognition [50]. Thermal imagery is considered as a natural presentation attack countermeasure, as it provides evidence of the user's liveness through simple acquisition. Generally, the fake artefact is characterized by thermal properties that are entirely different from those of a human face. Figure 2.7 reveals some examples of presentation attack in visible and thermal spectrum. It can be noted that simple presentation attacks, from (a) to (d), deliver thermal prints that are practically uniform. Although, when a silicone mask is worn by a person it can get heated and present a similar thermal print to a human face, as shown in Figure 2.7e, yet it delivers an average temperature much lower than of an average human face.

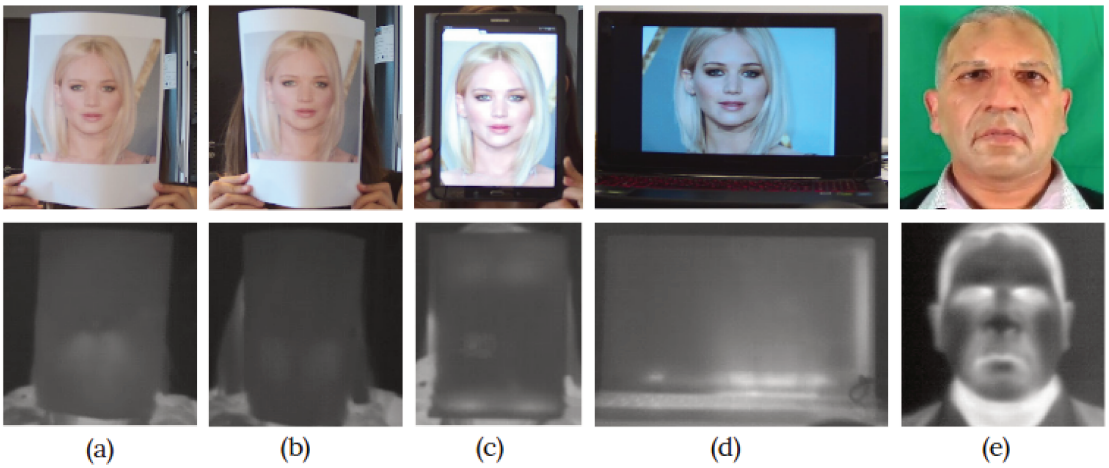


Figure 2.7: Presentation attack on visible and on thermal spectrum.(a) plain printed paper (b) wrapped printed paper (c) tablet (d) laptop (e) silicone mask (a sample from CSMAD database [51]).

Several multispectral databases [51, 52, 53] were proposed to study the robustness of different spectra against presentation attacks. Bhattacharjee et al. [51, 54] proposed to simply use mean brightness intensity of the thermal face region as presentation attack detection score. Despite of its simplicity, the proposed approach is proved to be efficient yielding to high detection accuracies. Agarwal et al. [53] introduced a new multispectral database of presentation attack and studied the robustness of visible, NIR and thermal spectra against these attacks. By evaluating several presentation attack detection approaches, thermal spectrum yielded to the highest performance proving to be the most robust compared to the other spectra. George et al. [52] proposed a multichannel convolutional neural network using a joint representation from multiple channels: depth maps, visible, NIR and thermal spectrum, which improved highly the classification accuracy of attacks from bona fides on account of the robustness of thermal spectrum to presentation attacks.

2.2.6 Disguise and cosmetic makeup

While images in visible spectrum can be easily altered by disguise and/or cosmetic makeup, thermal spectrum is less affected by these alterations due to the acquisition of thermal properties.

Disguise is generally acted using various artificial accessories that are marked by a different thermal signature than of a human face which can be easily detected on thermal images [55]. Dhamecha et al. [56] proposed a new database of face disguise in visible and thermal spectrum. Samples from the database are presented in Figure 2.8. The

authors also introduced a patch-based classifier for disguise detection. The proposed approach uses intensity and texture encoders to classify face patches in visible and thermal spectrum as biometric or non-biometric. The non-biometric patches are discarded and local binary pattern (LBP) based face recognition is performed on the biometric patches.

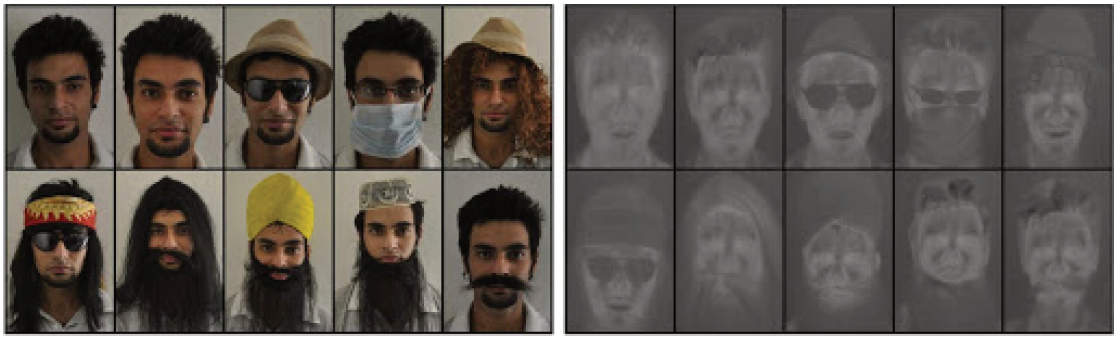


Figure 2.8: Samples of face disguise extracted from I^{2}BVSD database [56] in visible and on thermal spectrum.

Unlike disguise that can be easily perceived in thermal spectrum, cosmetic makeup hardly affects the thermal signature of a face, as can be seen in Figure 2.9. Therefore, face recognition can still be performed in thermal spectrum even in the presence of facial makeup changes. Short et al. [57] studied the impact of cosmetic makeup of different material on visible and on thermal spectrum. The authors conducted face recognition experiments in both spectra and proved that thermal spectrum has yielded to higher recognition performances than visible spectrum.

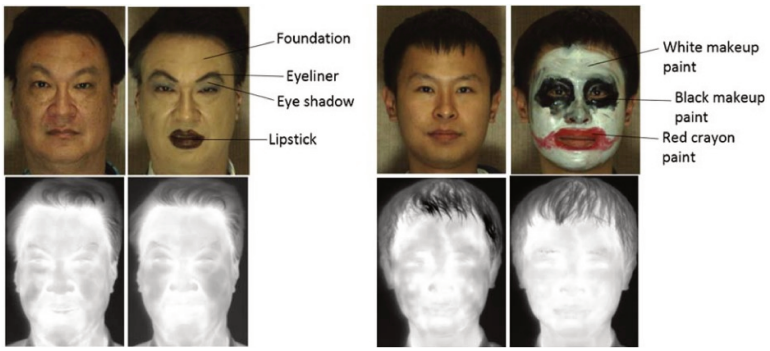


Figure 2.9: Impact of cosmetic makeup on visible and thermal spectrum. Figure extracted from [57].

2.2.7 Facial plastic surgery

It is acknowledged that facial plastic surgeries can alter the performance of face biometric systems operating in visible spectrum. Plastic surgery is generally used for correcting facial feature irregularities or improving facial appearance. This includes adding or

subtracting skin tissues, adding silicone, redistribute fat, etc. All these procedures require surgical incisions that cause alteration of blood vessel flow. These alterations are detectable in thermal spectrum as cold spots [55]. Figure 2.10 shows a thermal image of a leg where a cold spot appears indicating a surgical incision.

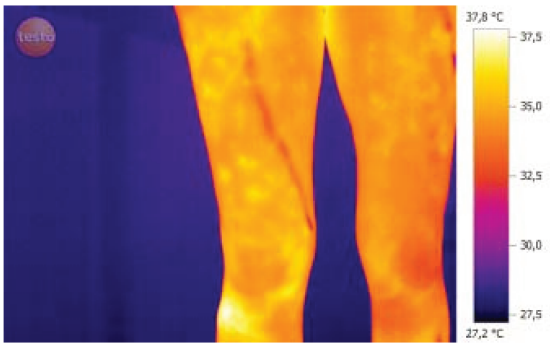


Figure 2.10: Thermal image showing pathological veins due to surgical incision. Figure extracted from [58].

2.2.8 Additional remarks

In addition to all the aforementioned advantages that the thermal spectrum grants over visible spectrum, thermal spectrum can differentiate monozygotic twins, while their appearance in visible spectrum is nearly identical. Prokoski et al. [59] carried out a qualitative assessment of similarity in visible and thermal spectrum using a limited number of samples acquired from monozygotic twins. The difference between the twins detected in thermal spectrum is traced back to the complexity of the network of blood vessels which provides a vascular pattern that is unique to each person including monozygotic twins.

While its listed advantages are numerous, thermal spectrum still suffers from various drawbacks, other than the eyeglasses problem. Thermal face images depend strongly on the heat pattern emitted by the face, however this emitted heat can be affected by a number of factors, such as ambient temperature, physical exercise, postprandial, illness, etc; as highlighted in [16]. Consumption of food, alcohol and caffeine may also alter the thermal characteristics. Some of these variables produce global changes to the face. But other variables affect the thermal appearance in a local manner, like blushing, or having a local infected area. This high sensitivity of the thermal images to several factors makes extracting discriminative features a difficult task.

Until very recently, thermal technology used to provide extremely expensive sensors with very low resolution. However these recent years, thermal technology is evolving rapidly offering competitive prices and higher quality sensors.

2.3 Summary

This chapter defines some background fundamentals of thermal imagery and the motivation behind its usage in facial image processing. In addition, a literature overview of facial image processing in thermal spectrum is presented for various unconstrained scenarios.

Chapter 3

Visible and thermal paired face database

Although thermal face recognition has recently grown as an active area of research, it still suffers from shortage of available thermal databases designed for training and evaluation of facial image processing that limits its exploration. In attempt to exploit the information complementarity provided by visible and thermal spectrum, a novel dual face database, that is acquired simultaneously in visible and thermal spectra, is introduced. The proposed database includes numerous facial variation such as expression, head pose, occlusion and illumination variations as to replicate the challenging scenarios encountered by face biometric systems. The remainder of the work presented in this dissertation are based on the database introduced in this chapter.

In this chapter, we introduce the first contribution of the work presented in this dissertation. Section 3.1 presents an overview of the existing public databases providing visible and thermal face images. Then, the proposed database that addresses the lack of variability in the existing ones, is introduced in Section 3.2 aiming to develop face recognition systems robust against real-world challenges. Section 3.3 presents a preliminary study conducted to assess the performance of the visible and the thermal spectrum under each variations. Following, a comparative study of different levels of fusion of visible and thermal spectra is conducted to conclude the saliency of each spectrum under different variations. Finally, a summary of the chapter is presented in Section 3.4

3.1 Overview of the existing visible and thermal face databases

Currently, there are numerous public face databases acquired in visible spectrum covering all variations possible [60]. However, interest in utilizing thermal face images has grown only recently and thus only a few databases have been provided, particularly databases that involve simultaneously acquired images in visible and thermal spectra. We present in the following the few public databases containing visible face images and their thermal counterpart. Table 3.1 summarizes the key descriptors of the presented databases.

3.1.1 EQUINOX

The "*human identification at a distance*" [61], collected by *Equinox Corp.*, is the most used database for evaluating face recognition algorithms based on thermal spectrum. The data was collected under 3 different lighting conditions (frontal, lateral right and lateral left), using a system composed of a visible CCD array and a LWIR microbolometer, capable of capturing simultaneous co-registered videos. During the acquisition, the subjects were asked to pronounce some vowels, and then to act out some expressions (smile, frowning, and surprised).

3.1.2 UND-X1

"*UND collection X1*" [62, 63, 64] is a thermal and visible facial database collected by the *University of Notre Dame*, using a Merlin uncooled LWIR sensor and a high resolution visible color camera. The data was acquired, in multiple sessions, under only two lighting conditions. For each illumination, two images were taken (neutral face and smiling).

3.1.3 USTC-NVIE

"*The natural visible and infrared facial expression database*" [65] was collected by the *University of Science and Technology of China*, using a DZ-GX25M visible camera and a SAT-HY6850 thermal camera. Each subject was asked to act out 6 different expressions, and then was exposed to situations provoking these expressions naturally and capture additional 6 different samples.

3.1.4 IRIS

IRIS thermal visible face database [66] is a public database collected by *Imaging, Robotics & Intelligent Systems Lab*. The data was acquired using a Panasonic WV-CP234 visible camera and a Raytheon Palm-IR-Pro thermal camera of 7-14μm spectral range. The database contains face images from 32 subjects asked to perform three different expressions. Five illumination conditions were considered. The two cameras are placed on a mechanized setup in a way that 11 images are captured from different viewing angles for each illumination and expression variation.

3.1.5 CARL

Carl Dataset [67, 68] is a public database collected by the *Polytechnical University of Catalonia*. The database contains face images from 41 different subjects in near-infrared, thermal, and visible spectrum. The data is acquired using a CMOS image sensor for visible spectrum and TESTO 880-3 thermal camera with spectral range of 8-14μm. Carl Dataset contains images from 41 subjects using 3 different illumination setups.

Table 3.1 sums up the key descriptors of the aforementioned databases.

Database	**Thermal resolution**	**#subjects/#images**	**# facial variations**			
			Illumination	**Expression**	**Head pose**	**Occlusion**
Equinox [61]	320×240	90/5000 pairs	3	3	1	1
UND-X1 [62, 63, 64]	312×239	32/2292 pairs	2	2	1	0
USTC-NVIE [65]	320×240	103/3230 pairs	1	2×6	1	0
IRIS [66]	320×240	30/2816 pairs	5	3	11	0
Carl [67, 68]	160×120	41/2460 pairs	3	1	1	0

Table 3.1: Existing face databases acquired in both visible and thermal spectra.

We should point out the fact that these databases were focused on different aspects of studies. The EQUINOX database [61] was collected in a single session, taking into account 3 expression variations and 3 light conditions. UND-X1 database [62, 63, 64] focused on studying time-lapse impact on thermal face recognition performance, the data was acquired in multiple sessions under two lighting conditions only, with neutral and smiling expressions. Whereas NVIE database [65] was acquired mainly to investigate the impact of thermal spectrum on expression recognition, thus the only variation considered was facial expression. The IRIS database [66] was designed to cover all the head pose variations. The CARL database [67, 68] was focused on studying multispectral face recognition under 3 different illumination conditions. UND-X1, USTC-NVIE, IRIS and CARL databases were collected using different devices to acquire face images in visible

and thermal spectra separately which does not guarantee the simultaneous acquisition resulting in not having the same face image in the two spectra. However, EQUINOX database [61] was collected using a sensor capable of capturing simultaneous videos in both domains. Among all the reviewed databases, the IRIS database seems to cover the widest range of facial variations. However, the visible and the thermal images are taken from different viewing angles. Lastly, although occlusion variations are still a challenging factor for face recognition algorithms, none of the databases have considered these variations.

3.2 Visible and thermal paired face database

The collection of a new database of visible face images and their thermal counterpart is motivated by the limited number of the public face databases providing paired images acquired simultaneously, and the lack of facial variations considered. In this section, we present the sensor used in the database collection, the acquisition setup, and a description of the collection protocol.

3.2.1 Dual Visible and thermal camera - FLIR Duo R

FLIR systems [69], acronym for forward-looking infrared, is the world's largest company specializing in thermal cameras and sensors production. A thermal imaging camera is a non-invasive instrument which scans and visualizes the temperature distribution of surfaces of an object rapidly and accurately.

The sensor used in collecting the database, presented in this section, is a newly (at the time of collection of the database) developed dual sensor thermal camera FLIR Duo R by FLIR Systems, featured in Figure 3.1. This camera is designed for unmanned aerial systems (UAS), but it is well suited for simultaneously capturing images and videos in both visible and thermal spectrum. The camera can be easily configured and operated using the FLIR UAS mobile application which allows to set color palettes, image optimization features and many other parameters shown in Figure 3.1. The visible sensor is a CCD sensor with a pixel resolution of 1920×1080. The thermal sensor of this camera is an uncooled Vanadium Oxide (VoX) microbolometer and has a spectral response range of 7.5 - 13.5μm with a pixel resolution of 160×120 and a noise equivalent temperature difference NETD<100mK. We acknowledge that the thermal resolution of the camera is considerably low. However, Mostafa et al. [70] has proven that high face recognition rates can be achieved with low resolution 64×64 pixels thermal face images, making of the camara's resolution a minor drawback. Moreover, an updated version of the camera with 640×512 resolution has been released later on, and a high resolution

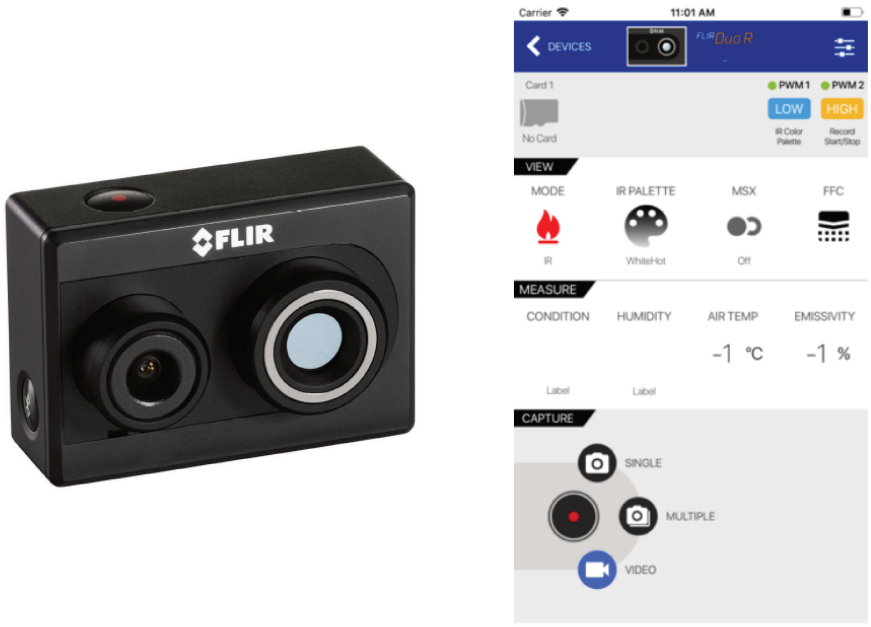


Figure 3.1: Flir Duo R camera and FLIR UAS mobile app

version of the database is being collected by this author of this dissertation, which will be shortly made available to the public.

3.2.2 Acquisition setup

The acquisition setup, illustrated in Figure 3.2, included a white background behind a chair at a fixed distance of 1m to the camera. The scene was illuminated with a three-point lighting kit, including a rim light, key light and fill light, placed to limit shadows. The ambient temperature of the room was set to 25°C. The room windows were covered with cardboard to achieve very low illumination conditions when illumination variations are acquired.

3.2.3 The database collection protocol

50 subjects of different age, sex and ethnicity volunteered for the collection of the database. The demographic characteristics of our proposed database are presented in Figure 3.3.

Before the acquisition process, volunteers were asked to fill and sign consent and metadata forms approved by the CNIL "*Commission nationale de l'informatique et des libertés*" [71]. During the data collection, the camera was set to capture a shot every second to limit acquisition errors. Each subject was asked to perform several facial

Figure 3.2: The database acquisition setup.

expressions, to change the head pose, to wear some items like sunglasses and cap, and finally the light was varied while the subject stayed in a natural state.

The database includes 21 face images per subject with different facial variations, resulting a total of 4200 images. The considered variations are shown in Figure 3.4 and described as follow:

- Expression: 7 pairs captured with standard illumination, frontal pose with different face expression: neutral, happy, angry, sad, surprised, blinking, yawning.
- Head pose: 4 pairs captured with standard illumination, neutral expression with different head poses: up, down, right at 30°, left at 30°.
- Occlusion: 5 pairs captured with standard illumination, frontal pose, neutral expression and varying occlusions: eyeglasses, sunglasses, cap, mouth occluded by hand, eye occluded by hand.
- Illumination: 5 pairs captured with frontal pose, neutral expression and different illuminations: ambient light, rim light, key light, fill light, all lights on, all lights off.

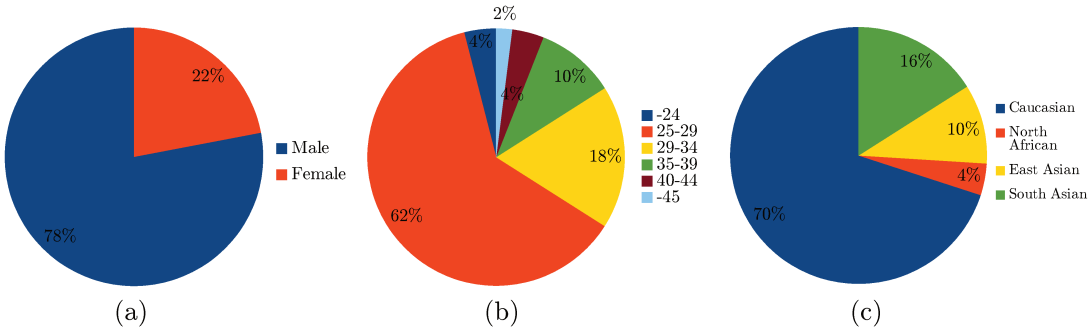


Figure 3.3: Demographics of VIS-TH database: (a) gender, (b) age, and (c) ethnicity.

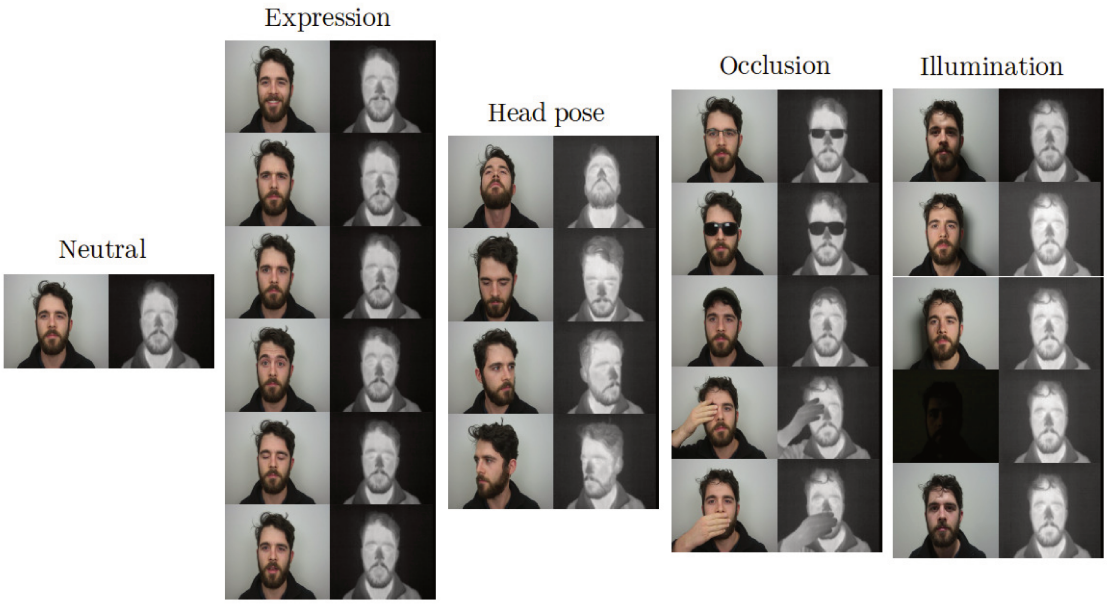


Figure 3.4: Illustration of visible and thermal images for various facial variations.

3.2.4 Access and usage conditions

The VIS-TH database proposed in this chapter is freely distributed upon request for standardization and academic research purposes, according to the European General Data Protection Regulation GDPR*. The database information and license request can be accessed at **http://vis-th.eurecom.fr/**.

*General Data Protection Regulation https://gdpr-info.eu/

3.3 Preliminary evaluation

The aim of this section is to present a preliminary evaluation to assess the applicability of the proposed database. A comparison of thermal and visible spectra against various facial variations introduced in our database is performed. This study will provide an efficient comparison of the performance of visible and thermal spectrum in the face recognition application, thanks to the simultaneous acquisition of the data that allowed to eliminate all other factors that may bias the comparison. Finally, a comparative study of different levels of fusion of visible and thermal spectra is carried out.

3.3.1 Evaluation protocol

Visible images were subsampled into 160×120 pixels. Faces in both visible and thermal spectra were detected and cropped. Face images were then normalized. Two benchmark approaches for face recognition were selected for our preliminary evaluation:

Eigenfaces [72] is a holistic approach based on principal component analysis (PCA). The idea of using principal components to represent human faces was developed by Sirovich and Kirby [73]. Eigenfaces approach is still considered as a baseline comparison method to demonstrate the minimum expected performance of a system.

Fisherfaces [74] is based on both principal component analysis (PCA) and linear discriminant analysis (LDA). Fisherfaces algorithm has achieved high performances on visible face images. Moreover, Socolinsky et al. [4] have compared holistic face recognition algorithms and proved that Fisherfaces achieved the highest recognition rate on thermal face images.

Performing a cross-fold validation, the data has to be split randomly in two subsets, one will be selected as a training set and the other as a testing set. Reiterating this process and returning the average performance reports significant results. However, since our aim is to study the impact of different variations on face recognition performance for visible and thermal face images, the database was split in 4 subsets, with each subset associated with a variation: illumination, expression, pose and occlusion. In order to test the face recognition performance for each variation, we have repeated the experiment considering, at each iteration, a different variation subset as training. For instance, to assess the face recognition performance on visible and on thermla spectrum under expression variation, the testing set, in this case, is the set containing images representing all the expression variations and the experiment will be repeated considering a different

training set at each iteration (illumination, pose and occlusion).

3.3.2 Face recognition in thermal and in visible spectrum

Table 3.2 and Table 3.3 illustrate the Rank-1 recognition rates of Eigenfaces (PCA) and Fisherfaces (LDA) algorithms on each spectrum. In addition, cumulative match characteristic (CMC) curves [75] are presented for visible and thermal spectra under different variations. A CMC curve shows various probabilities of recognizing a person depending on how similar their biometric features are to that of other people's. Figure 3.5 shows the overall CMC curves for Eigenfaces and Fisherfaces, representing results aggregated over visible and thermal spectrum. Each plot of Figure 3.5 represent CMC curves under different facial variation.

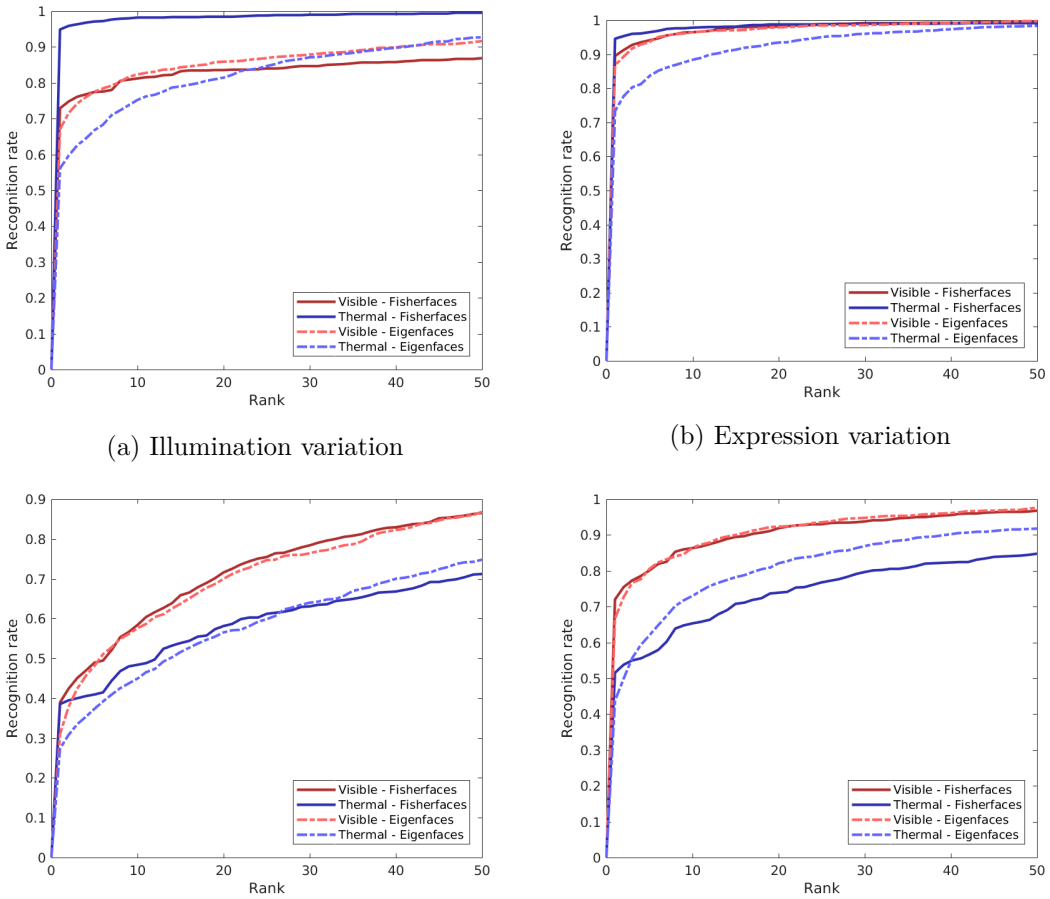
		TEST							
		Illumination				Expression			
		VIS		TH		VIS		TH	
		PCA	LDA	PCA	LDA	PCA	LDA	PCA	LDA
T	Illumination	N/A	N/A	N/A	N/A	0.703	0.814	0.606	0.96
R	Expression	0.857	0.733	0.765	0.973	N/A	N/A	N/A	N/A
A	Pose	0.854	0.66	0.708	0.893	0.617	0.914	0.446	0.891
I	Occlusion	0.891	0.793	0.725	0.973	0.69	0.957	0.63	0.962
N	**Average**	**0.867**	**0.728**	**0.733**	**0.946**	**0.67**	**0.895**	**0.56**	**0.937**

Table 3.2: Rank-1 recognition under expression and illumination variations.

		TEST							
		Pose				Occlusion			
		VIS		TH		VIS		TH	
		PCA	LDA	PCA	LDA	PCA	LDA	PCA	LDA
T	Illumination	0.352	0.312	0.284	0.365	0.706	0.69	0.45	0.59
R	Expression	0.296	0.476	0 .268	0.417	0.667	0.83	0.503	0.53
A	Pose	N/A	N/A	N/A	N/A	0.627	0.633	0.36	0.42
I	Occlusion	0.28	0.38	0.268	0.428	N/A	N/A	N/A	N/A
N	**Average**	**0.309**	**0.389**	**0.273**	**0.382**	**0.667**	**0.719**	**0.436**	**0.513**

Table 3.3: Rank-1 recognition under pose and occlusion variations.

As can be seen, thermal spectrum outperforms the visible spectrum when tested on the illumination variation. This confirms the statement that thermal spectrum does not need an external source of illumination to acquire images while visible spectrum is highly sensitive to light changes. Similarly when tested on expression variation, we note that face recognition performance is particularly higher for the thermal spectrum compared to visible spectrum. We believe that this outcome is due to the reflective nature of visible spectrum that makes it highly sensitive to light changes unlike the thermal spectrum,



(a) Illumination variation

(b) Expression variation

(c) Head pose variation

(d) Occlusion variation

Figure 3.5: Cumulative Match Characteristic curves for various collection scenarios.

since changes in facial expressions imply changes in the distribution of the light across the face surface. Although, when it comes to head pose variations, we notice that both visible and thermal spectrum perform almost equally at Rank-1 recognition. Furthermore, performance obtained by visible spectrum is significantly higher than the performance of thermal spectrum for occlusion variation. This is due to some limitations of the thermal spectrum. For example, the eye glasses are opaque to the thermal wavelengths since they block the heat emitted by the face region covered by the glasses' frame and lenses, while on visible spectrum we can see the eye details thanks to visible light transmittance in glass. Comparing the face recognition performance obtained using the two benchmark face recognition algorithms, we observe that the performance of the Fisherfaces approach on thermal spectrum is significantly higher than the performance of the Eigenfaces method, exclusively for illumination variation (Figure 3.5a) and to a lower degree for expression variations (Figure 3.5b). However, this increase in performance is not observed for visible spectrum. This is justified by the fact that intra-class variability in thermal

spectrum is considerably smaller than intra-class variability in visible spectrum. Light distribution across the face changes according to the illumination conditions and to some extent to the expression conditions, leading to a high variability in visible images but not in thermal images as the thermal spectrum is immune to light changes.

3.3.3 Comparative study of different levels of fusion

In this section, we present early experiments in sensor-level, feature-level and score-level fusion to study the impact of different levels of fusion on face recognition rate on the proposed database and to infer the saliency of each spectrum against each variation.

Preprocessing

One of the main challenges of sensor-level fusion is that it requires high precision in image registration. The data acquired with the new sensor FLIR Duo R presents a slight shift. Visible and thermal face images were co-registered using edge-based image registration approach inspired from [76].

Schemes of different levels of fusion

In the sensor-level fusion approach, pixels values of visible and thermal images are weighted and summed to generate fused images. Face recognition experiments are then performed on the fused face images. Figure 3.6 illustrates a fused image 3.6c resulting from the average summation of Figure 3.6a and Figure 3.6b. We can observe that the fused image presents the properties of both visible and thermal spectra.

(a) Visible image (b) Thermal image (c) Fused image

Figure 3.6: Sensor-level fusion of visible and thermal spectra.

For feature-level fusion, we compute separately the face subspace from the training set for each of the spectra. For testing set, the projection of gallery and probe faces are done onto the corresponding face subspace. Visible features and thermal features are

then fused through weighted summation.

Whereas for score-level fusion, face subspaces are computed separately for the visible images and its thermal counterpart. Scores, for the visible and the thermal spectra, are then computed between the gallery and the probe faces. Then, these scores are normalized, using min-max normalization. Finally, the scores are fused using a weighted summation.

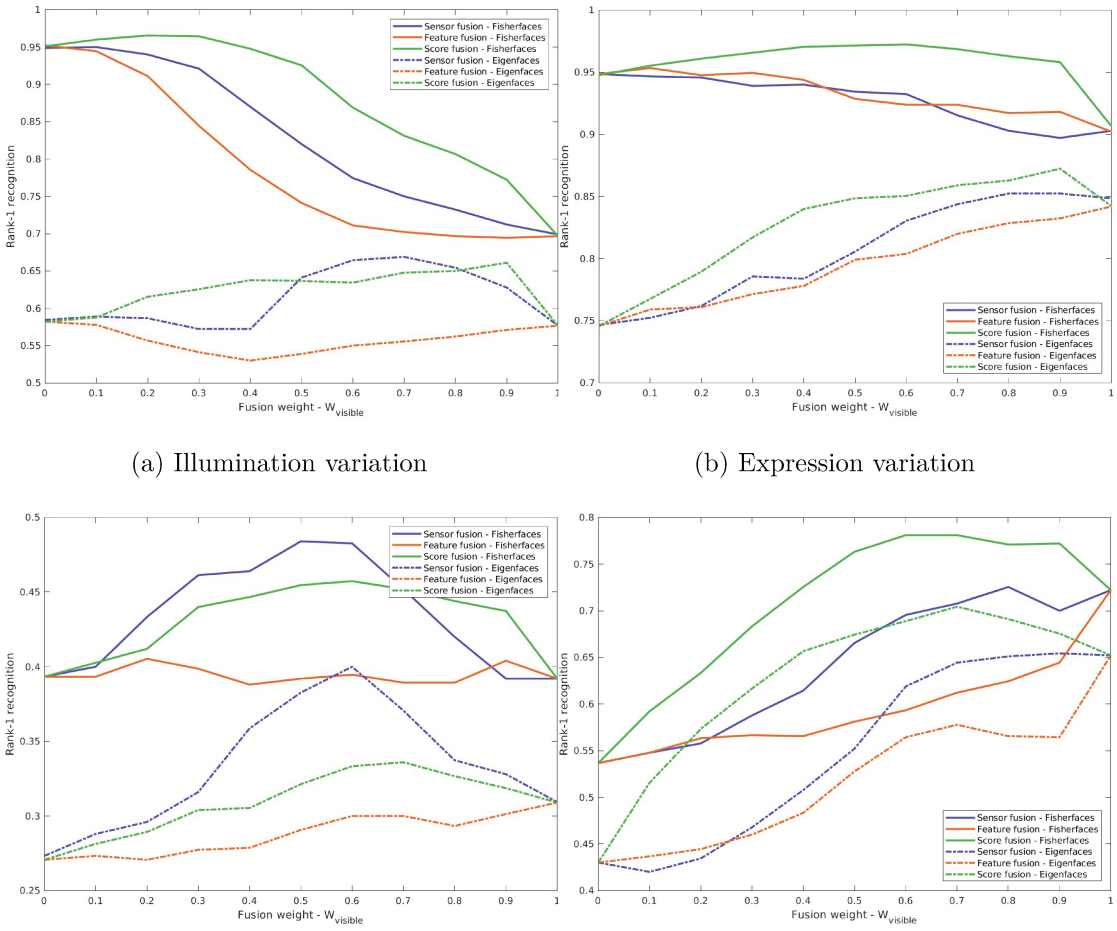
For the three proposed schemes of fusion, we have varied the weight associated with the visible spectrum as well as the thermal spectrum, as illustrated in Equation 3.1 where $fused$, $visible$ and $thermal$ refer to either the image, the face feature or the matching score and computed rank-1 recognition for Eigenfaces and Fisherfaces algorithms for each weight.

$$fused = W_{visible} \times visible + (1 - W_{thermal}) \times thermal \tag{3.1}$$

Experimental results

To study the impact of different levels of fusion on face recognition performance for each spectrum, we present, in Figure 3.7, the variation of recognition rate according to the weight associated to the visible and the thermal spectrum when tested under different facial variations.

For illumination variation, it is already proved that face recognition systems based on thermal spectrum perform better than the system based on visible spectrum. However, in particular for sensor-level fusion, we have perceived when we have added the visible information the recognition rate has relatively increased and that is due to the textural information that the visible spectrum provides. Although after a certain threshold, the more visible information we consider, the more the performance decreases. This observation can be justified by the fact that visible spectrum is highly sensitive to illumination changes. Figure 3.7b illustrates the impact of fusion levels on recognition rate under expression variation. We observe that score-level fusion provides the highest performance rates. However, the performance has hardly increased compared to the performance of thermal based face recognition. Considering now the recognition performance when tested under head pose variation featured in Figure 3.7c, it is noted that the performance has drastically increased when the two spectra were uniformly fused. Particularly, the highest performance rates were registered when sensor fusion was applied. We believe this improvement is due to the combination in image level of the textural information of the visible spectrum and the invariance of thermal spectrum to light distribution across



(a) Illumination variation

(b) Expression variation

(c) Head pose variation

(d) Occlusion variation

Figure 3.7: Impact of different fusion levels on the rank-1 recognition rate varying the weight associated to each spectrum.

the face. Whereas for occlusion variation, a reverse behaviour is observed when compared to illumination variation. The poor performance of thermal spectrum when tested under occlusion variation is due to the fact that certain objects, i.e occlusions, block the heat emission.

Overall, when a spectrum performs considerably higher than the other, we did not obtain significant improvement in performance of face recognition when applying fusion. Also, it is perceptible that score-level fusion provides better results than sensor-level fusion. This observation can be justified by the fact that fusing images from different spectra can result in altering the information provided by each, in particular when one spectrum fails under specific conditions, as it is the case of low illumination for visible spectrum and eyeglasses for thermal spectrum.

3.4 Summary

A new database of face images acquired simultaneously in thermal and visible spectra, aiming to cover a wider range of facial variations compliant with hands-on scenarios, is introduced in this chapter. The proposed database is publicly available* upon request. Preliminary evaluation is presented to assess the applicability of the proposed database to the face recognition task and to determine the performance of state-of-the-art benchmark face recognition approaches for both visible and thermal spectrum. In addition, a comparative study of different fusion levels was conducted to gauge the saliency of each spectrum in improving face recognition performance.

*VIS-TH database: http://vis-th.eurecom.fr/

Chapter 4

Cross-spectrum face recognition based on thermal-to-visible image synthesis

Face synthesis from thermal to visible spectrum is fundamental to perform cross-spectrum face recognition as it simplifies the integration of thermal technology in already deployed face recognition systems and enables manual face verification. In this chapter, a new solution based on cascaded refinement networks is proposed. This method generates synthesized visible images of high visual quality without requiring large amounts of training data. By employing a contextual loss function during training, the proposed network is inherently scale and rotation invariant. We discuss the visual perception, followed by a qualitative evaluation of the synthesized visible faces in comparison with recent works. We also provide an evaluation in terms of cross-spectrum face recognition, where the synthesized faces are compared against a gallery in visible spectrum using two state-of-the-art deep learning based face recognition algorithms. The evaluation results show the efficiency of the proposed approach and pave the way to its exploration for further facial image processing tasks.

The remainder of this chapter is organized as follow. Motivation that drove to this work are presented in Section 4.1. The proposed approach for thermal-to-visible image synthesis is introduced in Section 4.2. Section 4.4 details the adopted experimental setup. A qualitative and quantitative assessment of the synthesized visible images is presented in Section 4.5. Following, an evaluation of the proposed approach in terms of cross-spectrum face recognition is reported in Section 4.6. The chapter is summarized in Section 4.7.

4.1 Context and motivation

While thermal face processing [15, 45, 77, 78, 79] has evolved during the last two decades, the deployment of thermal technology remains a step behind compared to technologies deployed in visible light spectrum. The motivation behind the work presented in this chapter relates to the need of a prompt and straightforward integration of thermal sensors in already deployed face recognition systems. However, enrollment data of these existing systems are commonly acquired exclusively in visible light spectrum. Recollection of enrollment samples in thermal spectrum would be costly in terms of time, efforts, and financial and storage resources, and is thus an un-realistic alternative to thermal face recognition deployment. Many studies [27, 80, 81, 82, 83, 84, 85] have attempted to match thermal face images against visible face enrollment samples. Considering the large difference between the visible and the thermal spectra, several efforts have been made to try to overcome this gap. These can be categorized into three aspects: latent subspace, domain invariant features and image synthesis.

Latent subspace approaches aim to project faces acquired in both spectra into one common underlying subspace, in which the relevance of thermal-to-visible data can be directly measured. Choi et al. [82] [27] used Partial Least Squares Discriminant Analysis (PLS-DA) to learn the mapping between thermal and visible face images. Safraz et al. [80] used a multilayer fully-connected feed-forward neural network to learn the non-linear mapping between the two modalities over the training set while preserving the identity information. The second approach to perform cross-spectrum face recognition seeks to extract domain invariant features, that are only related to face identity. Chen et al. [83] introduced a thermal-to-visible matching framework based on hidden factor analysis used to extract the identity features of a person across different spectra. Image synthesis approaches aim to convert a face image from one spectrum to another, so that face matching can be carried out in the same domain. In this work, we focus on an image synthesis strategy for cross-spectrum face recognition, consisting in generating visible images from thermal captures that will be matched against a gallery of visible faces. This approach bridges the spectrum gap at the image preprocessing, as illustrated in Figure 4.1, without requiring modification on inner modules of the face recognition system. Opting for this strategy is essential to enabling the integration of thermal face data in existing face recognition systems, as well as manual face verification by human examiners.

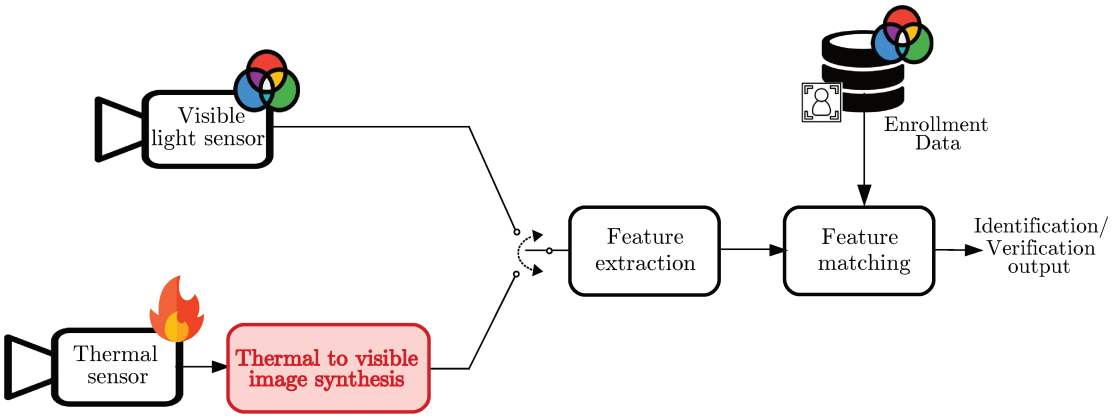


Figure 4.1: Illustration of image synthesis based cross-spectrum thermal-to-visible face recognition. In this case the integration of thermal technology in already deployed face recognition systems only requires the addition of a thermal-to-visible image synthesis module.

4.2 Literature overview

First attempts to investigate face synthesis from thermal to visible spectrum were conducted by Li et al. [81]. Their work presented a learning-based framework that takes advantage of the local linearity in the spatial domain of the image as well as in the image manifolds. Then, they apply Markov random fields to organize the image patches and improve the estimated visible face images. Dou et al. [86] used Canonical Correlation Analysis (CCA) to extract the features in order to find one-to-one mapping between thermal and visible faces. The relationship between the two feature spaces in which the visible features are inferred from the corresponding thermal features is then learnt using locally linear regression. Finally, locally linear embedding is utilized to reconstruct the visible face from the converted thermal features.

In the wake of the recent advances in deep learning, several works were based on Generative Adversarial Neural networks (GAN) to synthesize visible images not only from thermal inputs [87, 88], but also from near-infrared [89, 90], and polarimetric data [84, 91]. GANs, first introduced by I. Goodefellow in [92], can learn to generate from any distribution of data through a contest of two neural networks: a generator and a discriminator. The generator aims to maximize the probability of making the discriminator classify its output as real. While the discriminator pushes the generator to generate more realistic data.

Different models can also be used for similar conversion. For examples, deep con-

volutional Generative Adversarial network (DCGAN) [93] and Boundary Equilibrium Generative Adversarial Networks (BEGAN) [94]. DCGAN introduced the Convolution Neural Network (CNN) into the discriminator and the generator. BEGAN introduced an equilibrium factor that controls the model training by balancing the discriminator and generator. These GAN models significantly improved the training stability, but they did not improve the generated images quality. However, and notwithstanding the more complex resulting topologies, some GAN-based approaches such as Cycle-Consistent Adversarial Networks (CycleGAN) [95] and Image-to-Image Translation with Conditional Adversarial Nets (Pix2Pix) [96] succeeded at generating higher resolution images. CycleGAN consists of four neural networks (two generators and two discriminators). Training such a big model is computationally costly and requires large databases, that are unavailable for an application like the one dealt with in this chapter, to achieve satisfactory results.

Zhang et al. [84] considered synthesizing colored faces from thermal images with various head poses and occlusion with eyeglasses. This work used Conditional GANs inspired from the Pix2Pix system [96], but coupled with a closed-set face recognition loss that led to preserve the face identity information. A cross-spectrum face recognition evaluation is performed, using the pre-trained MatConvNet VGG-based model [97], and reported a performance improvement of 14.88% compared to the Pix2Pix [96] system's reported performance. A recent work by Wang et al. [88] derived from the CycleGAN model [95] incorporated a facial landmark detector loss that depicts face identity preserving features. This system was evaluated using a FaceNet model [98] pre-trained on publicly available visible datasets, and improved cross-spectrum face recognition performance by 3% compared to the original CycleGAN system. However, this work is different from our framework in that its aim is to generate visible face images in gray scale, and it also discarded face generation under challenging conditions such as head pose and occlusion.

4.3 Thermal-to-visible image synthesis

To generate images from thermal to visible spectrum, we propose to base our approach on cascaded refinement networks (CRNs) [21]. We chose the CRN as the basic block for our image synthesis as it considers multi-scale information and is based on training a limited number of parameters. This allows for a higher resolution generation and less data size dependency in comparison to solutions based on GANs. Chen at al. [21] have adopted pixel-to-pixel loss, perceptual loss [99], to train the CRN model. We, on the other hand, used contextual loss [100], that compare regions of images based on semantic meaning. In this section, we first present the CRN network architecture. Then, we introduce the

contextual loss function.

4.3.1 Cascaded refinement network

Cascaded refinement network was first presented in [21] to synthesize photographic images from semantic layouts. The presented architecture scales seamlessly to high-resolution images, obtaining 2-mega pixels photo-realistic images from 2D semantic label maps. The challenge addressed in [21] lies in the attempt to generate detailed photographic images from simple semantic label maps. Thermal-to-visible image synthesis can be seen as a similar problem to the one dealt with in [21], as our objective is to generate highly informative images in visible spectrum from a less informative domain as thermal spectrum, since it lacks texture and color information.

CRN is a feed-forward convolutional neural network that consists of inter-connected refinement modules. The first module considers the lowest resolution space (4×4 in our case) and takes as input the thermal image downsampled to 4×4. A feature map is generated by the first module and then upscaled using a simple bilinear upsampling. The next module receives as input the upscaled feature map concatenated with the thermal image downsampled to the corresponding resolution. Image resolution is duplicated in the successive modules until the last module (128x128 in our case), matching the target image resolution. An illustration of the image synthesis approach using CRN is shown in Figure 4.2. The input thermal images are processed at different scales and fed into the next level in the cascade along with the thermal image at the next scale. Finally, the targeted image (visible in this case) is synthesized. Figure 4.3 portrays a single refinement module. Each refinement module consists of only three layers, input, intermediate, and output layer, and handles a given resolution. Global structure of visible face is generated at low resolutions while local details are progressively refined.

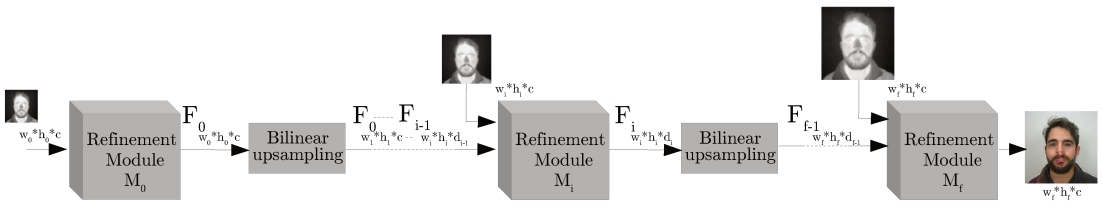


Figure 4.2: The CRN-based multi-scale approach to transform the thermal image into a visible image.

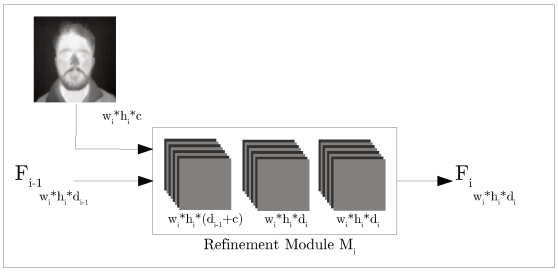


Figure 4.3: Illustration of the refinement module. As an input, the refinement module gets the feature map generated by the previous module concatenated with the thermal image downscaled at the corresponding resolution $w_i \times h_i \times c$.

4.3.2 Contextual loss

To control the training of our CRN network, we used the contextual loss function [100]. This choice is based on our need for: a) a loss function that is robust to not well aligned (as in our use-case where input face images are not uniformly aligned), and b) neglect outliers on the pixel level (in comparison to pixel-level loss [96, 101]). Gramm loss [102] can satisfy the two aforementioned conditions, however, unlike in the contextual loss, it does not constrain the content of the generated image as it describes the image globally.

The contextual loss function aims to compare regions with similar semantic details while preserving the context of the entire image. Contextual loss is based on contextual similarity measure. Two images are considered contextually similar if their corresponding sets of features are similar. Figure 4.4 presents a simplified illustration of the idea behind measuring contextual similarity. The feature y_j is contextually similar to feature x_i if the distance between the two features is particularly small compared to the rest of features in image X. This problem can be posed as nearest neighbor search in image X for each feature y_j. Put differently, contextual similarity is high when there is one-to-one matching of feature sets, while it is low when for a feature y_j it exists a set of features x_i that are almost equally similar to y_j. Accordingly, features x_i and y_j are contextually similar if $d_{ij} \ll d_{kj}$, while $\forall k \neq i$ and d_{ij} denotes the Cosine distance between features x_i and y_j. To highlight the similarity of features x_i and y_j in comparison to the other features x_k, distances are normalized as follow:

$$\tilde{d_{ij}} = \frac{d_{ij}}{min_k d_{kj} + \epsilon} \tag{4.1}$$

where $\epsilon = 1e - 5$. Distances are converted into similarity as:

$$w_{ij} = e^{\left(\frac{1-\tilde{d_{ij}}}{h}\right)} \tag{4.2}$$

where $h > 0$ denotes the bandwidth parameter. A normalization of the contextual similarity is then applied so that it becomes robust to scale variation:

$$CX_{ij} = \frac{w_{ij}}{\sum_k w_{kj}} \tag{4.3}$$

Finally, the contextual similarity between images X and Y, given N feature points, is formulated as:

$$CX(X,Y) = \frac{1}{N}\sum_i \max_j CX_{ij} \tag{4.4}$$

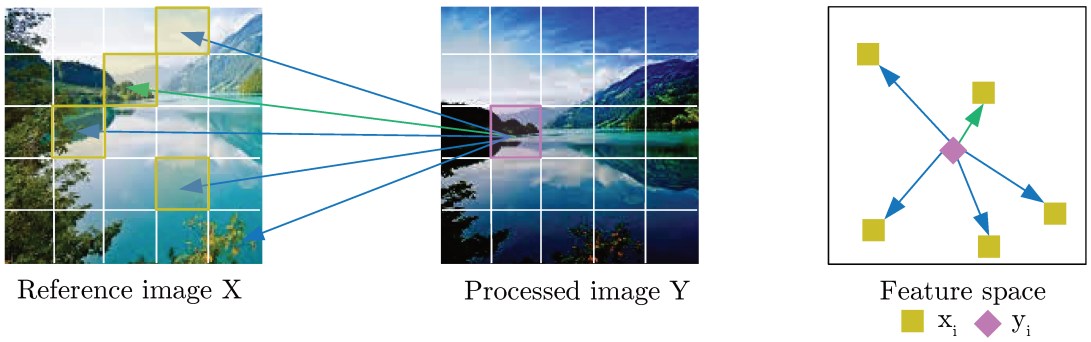


Figure 4.4: Illustration of contextual similarity. The patches of image Y are compared against all patches of image X at high dimensional space. The feature patch x_i in image X that corresponds to the feature patch y_j in image Y is presented at a closer distance in feature space compared to the other features from image X. This means the contextual similarity between the two features, linked with the green arrow, is higher that the contextual similarity between the rest of the sets of features, linked with the blue arrows.

The loss function of thermal-to-visible image synthesis should be able to transform the image from thermal to visible spectrum while preserving the facial attributes. Our loss function can then be modeled as a combination of two losses: style loss and content loss, as defined by Gatys et. al [102]. The style loss is computed between the synthesized visible image and the ground truth visible image. Minimizing the style loss manages to generate artificial images with the same properties as the target visible image. The content loss is computed between the input thermal image and the synthesized visible image. The content loss aims at preserving details of facial attributes. Using contextual

loss allows to tolerate some local deformations that are required to perform the thermal-to-visible style transferring. Both losses were calculated between image embeddings extracted by a pre-trained VGG19 [103] network trained on the ImageNet database [104]. The total loss is calculated as given in [100] and formulated as:

$$\begin{aligned} L_{CX}(I_{TH}, I_{VIS}, G) =& \lambda_1(-\log(CX(\Phi^{l_s}(G(I_{TH})), \Phi^{l_s}(I_{VIS})))) + \\ & \lambda_2(-\log(CX(\Phi^{l_c}(G(I_{TH})), \Phi^{l_c}(I_{TH})))), \end{aligned} \tag{4.5}$$

where I_{TH}, I_{VIS}, and G are the input thermal image, reference visible image, and the generator (i.e. thermal-to-visible image synthesis module) respectively. CX is the rotation and scale invariant contextual similarity [100]. Φ is a perceptual network, VGG19 in our work. $\Phi^{l_c}(x)$, $\Phi^{l_s}(x)$ are the embeddings vectors extracted from the image x at layer l_c and l_s of the perceptual network respectively. Here l_c is the `conv4_2` layer representing the content layer and l_s is the `conv3_2` and `conv4_2` layers representing the style layers, as motivated in [102]. Feature sets are considered as 5×5 patches extracted with stride of 2 from the content and style layers.

4.4 Experimental setup

In this section, we present the preprocessing steps we applied on the database used for the development and the evaluation of our proposed solution. Then, we introduce our implementation details set to perform thermal-to-visible image synthesis. Finally, we present the baselines models of image synthesis to which our approach is compared.

4.4.1 Database preprocessing

We used the VIS-TH face database [105], presented in chapter 3, for the development and the evaluation of our solution. As stated, pixel resolution of face images in visible spectrum is 1920×1080 pixels and in thermal spectrum is 160×120 pixels. Images, from both visible and thermal spectrum, were normalized and sampled to 128×128. Enabling an evaluation of our solution in hands-on scenarios, and considering that face alignment in thermal spectrum still remains a challenge itself, the face images were not aligned, thus they contained slight variable shifts.

4.4.2 Implementation details

In our implementation, the training was run for 40 epochs, batch size of one, and 1e-4 learning rate. The weights assigned to each term of the loss function are set to $\lambda_1 = 0.01$ and $\lambda_2 = 0.99$ by checking the resulting synthesized image visually. Moreover, the pairs of

input thermal image and reference visible images are of identical faces that are acquired simultaneously, and thus the loss weighted by λ_2 maintains the structural details of the source image.

Face images from 45 subjects, except for the ones acquired in total darkness, were used for training the face synthesis network. The thermal face images from the remaining 5 subjects were fed to the trained model to synthesize the visible images. This experiment was performed 10 times in order to cover all the images contained in the database without overlapping the test and train images or identities.

4.4.3 Image synthesis baselines

In order to assess the efficiency of our proposed approach to perform thermal-to-visible image synthesis, we have selected two baseline models. The two selected baselines are based on GANs, as it is the most used generative model since it was introduced in 2014 by Goodfellow et al [92]. The first baseline is the renowned Pix2Pix model, proposed by Isola et al. [96] to perform image to image translation. The second baseline is TV-GAN model presented by Zhang et al. [84]. This baseline is more adapted to our framework where the proposed model aims to synthesize visible face images from thermal inputs.

Isola et al. [96] , referred to as Pix2Pix, learns the mapping from one domain to another, by training a conditional GAN using a least absolute deviations (L1) loss function. The generator is based on the U-Net [106] architecture, an encoder-decoder with skipped connections between mirrored layers in the encoder and decoder stacks. At the same time, the discriminator aims to classify real images from generated ones. Pix2Pix model has been extensively used for a variety of tasks and applications. The training was run for 85 epochs, batch size of one, and 2e-4 learning rate.

Zhang et al. [84] , have designed a network, called TV-GAN, notably to generate visible face images from thermal captures. This work is inspired from Pix2Pix [96], as it uses the same exact network for the generator. However, the authors proposed a multi-task discriminator, that does not only classify real from generated images, but also performs a closed-set face recognition with which they can compute an identity loss. This aims to generate visible images while preserving identity information from the thermal inputs. The introduction of identity loss in the GAN training was inspired by Tran et al. [107]. The training was run for 65 epochs, batch size of one, and 2e-4 learning rate.

4.5 Quality assessment of synthesized visible images

The human visual cortex is exclusively trained on scenes which spans visible light wavelength detected by the human eye, much similarly to existing face recognition systems. Consequently, humans present very limited ability to interpret thermal images. The motivation behind thermal-to-visible image synthesis is not only limited to perform cross-spectrum face recognition, it is also driven by the need to convert images from thermal to visible spectrum so that it can be interpreted by humans. Visual quality assessment is then necessary. In this section, we present firstly a qualitative assessment of synthesized visible images. Then, a quantitative evaluation is reported by comparing the synthesized images to the reference visible images.

4.5.1 Qualitative assessment

The images in Figure 5.1 illustrate, in each row, a sample from different facial variations of synthesized visible face images from thermal inputs. The column (a) shows the input thermal faces. In columns (b) to (d), we present visible faces synthesized using the Pix2Pix model by Isola et al. [96], the TV-GAN model by Zhang et al. [84] and finally our model based on cascaded refinement network, respectively. The last column (e) shows the ground truth visible faces.

The different face images with frontal face pose were synthesized with satisfying visual quality. Although we note that our proposed model has succeeded in generating more informative details (e.g. eyes, mouth) compared to the Pix2Pix and TV-GAN results, it does not always generate the correct attributes such as skin color and gender. We can observe that all synthesized visible faces differ in skin color from the ground-truth images, and this applies to all synthesis models. This is due to the fact that thermal images do not contain texture and color information, thus, it is difficult to infer the skin color tone from the thermal signatures. Another visual distortion can be noted on the visible samples synthesized by our proposed model in the second and the fourth row of Figure 5.1. These samples show some added facial hair around the mouth and the jaw area. This observation can be reasoned by the unbalanced distribution of gender representation within the training data. Third and sixth rows display samples from different head poses, where we can observe major artefacts in the synthesized visible faces when compared to the frontal head pose. As for images acquired with occlusion, illustrated in the fourth and seventh rows, they were synthesized in relative good quality. However, we perceive some confusion in generating faces with eyeglasses. This is justified by the fact that the training data contains samples with eyeglasses and others with sunglasses that both have similar thermal pattern, both blocking the heat emitted by the

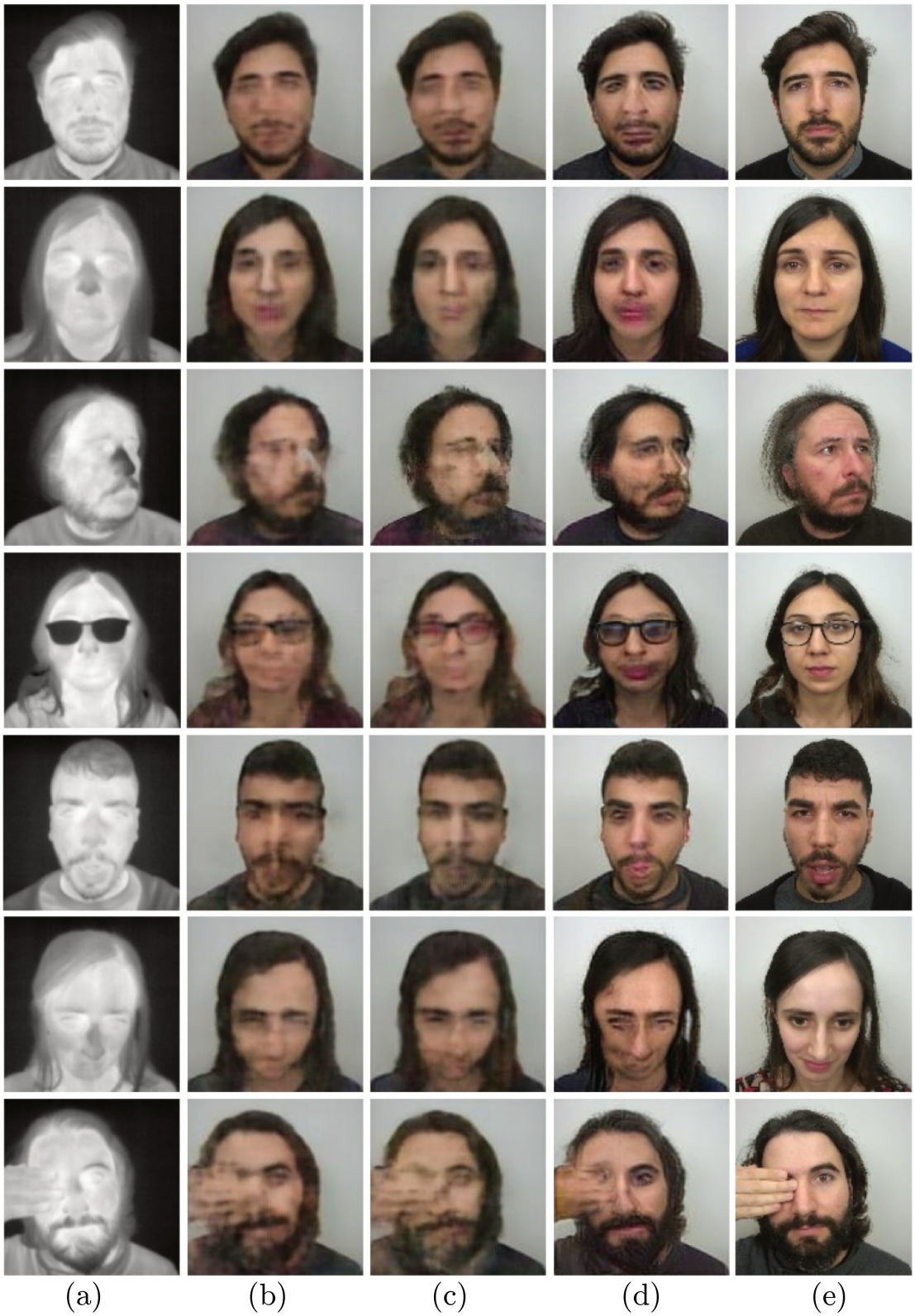


Figure 4.5: Selected samples of synthesized face images under challenging scenarios. **(a)** Thermal **(b)** Isola et al. [96] **(c)** Zhang et al. [84] **(d)** Ours **(e)** Ground truth

eyes area. Synthesizing visible images with occlusion by hand was successful, however, with high level of blur in the hand region. Overall, it is noteworthy that our proposed model provides visible faces that are the most visually pleasing compared to Pix2Pix and TV-GAN models.

To highlight the main motivation of this work, we display, in Figure 4.6, samples that were acquired in operative scenarios of thermal sensors usage, where face images were captured in total darkness. As expected, the poor or absent illumination does not impact the synthesized visible images. In fact, we succeeded in synthesizing images with informative facial attributes that are absent in the visible spectrum.

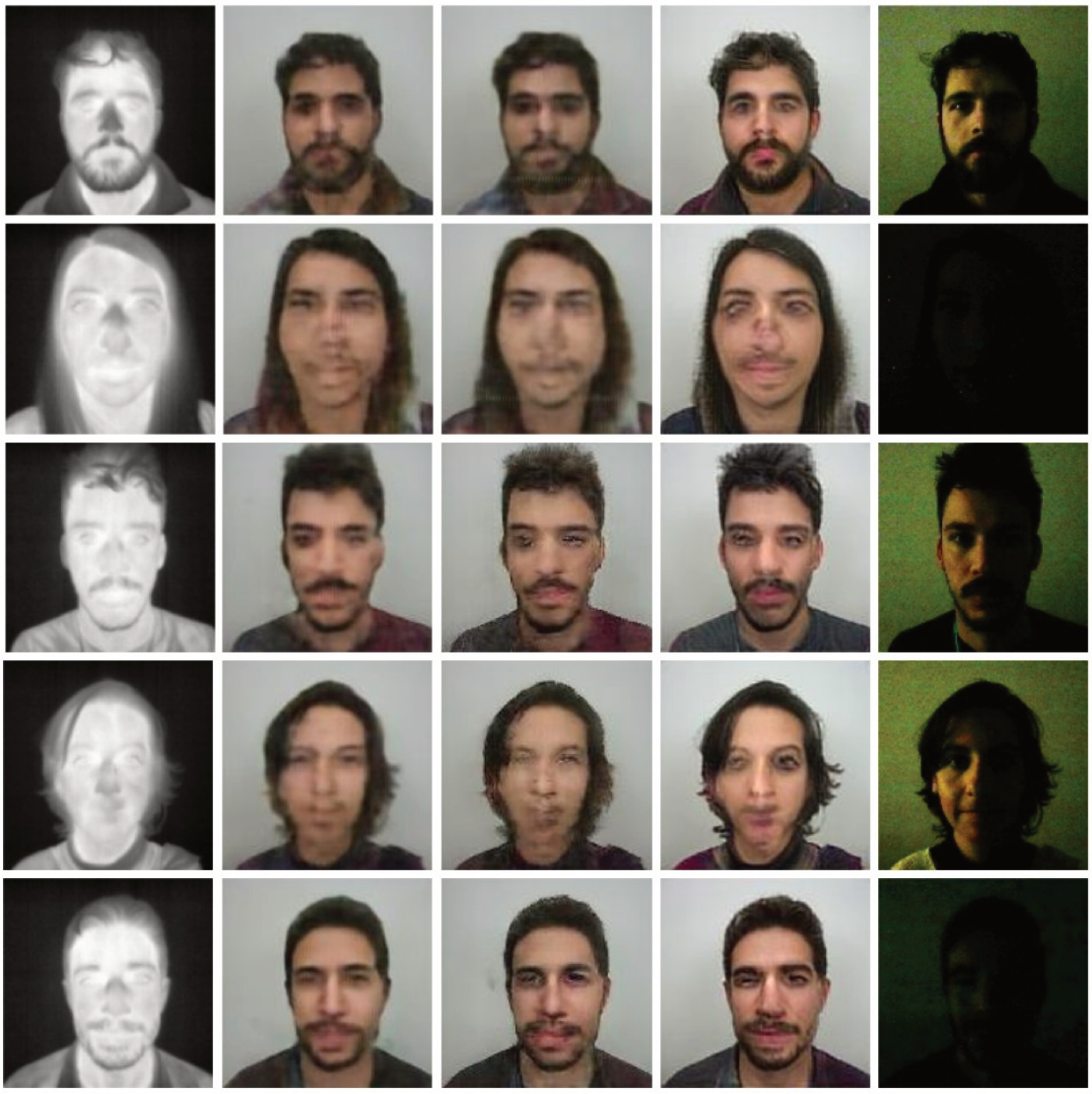


Figure 4.6: Samples of generated images acquired in total darkness. **(a)** Thermal **(b)** Isola et al. [96] **(c)** Zhang et al. [84] **(d)** Ours **(e)** Ground truth

4.5.2 Quantitative assessment

Two quality indices, peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM), are selected to assess the visual quality of the synthesized visible images.

Peak signal-to-noise ratio (PSNR) measures the level of degradation of a reconstructed signal in comparison to the original signal. A higher PSNR value indicates

higher image quality.

$$PSNR = 10 \log_{10} \left(\frac{\max_I^2}{MSE} \right)$$
$$MSE(I_{VIS}, G(I_{TH})) = \frac{1}{m \times n} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} (G(I_{TH})(i,j) - I_{VIS}(i,j))^2 \tag{4.6}$$

where MSE is mean square error, and $\max_I$ is the maximum pixel value of the image (255 for 8 bits images). I_{VIS}, I_{TH}, G indicate the reference visible image, the input thermal image and the image synthesis model, respectively. $G(I_{TH})$ represent the synthesized visible face image.

Structure similarity index measure (SSIM) was introduced by Wang et al. [108]. This quality metric is considered more adapted to the human visual system. SSIM measures the image degradation as the perceived alteration of the structural information. Let us suppose that x and y are two windows extracted from the reference visible face image I_{VIS} and the synthesized visible image $G(I_{TH})$, respectively. SSIM is then formulated as:

$$SSIM(x,y) = \frac{(2\mu_x\mu_y)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{4.7}$$

where μ_x and μ_y are the average of x and y, σ_x^2 and σ_y^2 are the variance of x and y, respectively. c_1 and c_2 are positive constant to prevent a null denominator.

Table 7.1 reports the PSNR and SSIM values obtained when comparing the synthesized visible face images, generated using different image synthesis models, to the ground truth visible images. The obtained results, ∼17dB for PSNR and ∼0.65 for SSIM, do not reflect high fidelity of the synthesized visible images to the ground truth. The synthesized visible faces are generated from facial thermal signatures, that represent different information. Thermal-to-visible image synthesis models aim to reproduce an estimation of visible light spectrum properties but it cannot predict them accurately, such as texture, color and more detailed geometrical information.

Comparing the results obtained for the two baselines, the identity loss term that was introduced by Zhang et al. [84], to the model proposed by Isola et al. [96] has led to a

slight increase of quality indices. However, a more relatively important improvement is noted for our proposed model, which aligns with our qualitative assessment of the image synthesis quality.

	PSNR (dB)	**SSIM**
Isola et al. [96]	17.247 (± 2.855)	0.6485 (±0.123)
Zhang et al. [84]	17.257 (±2.897)	0.6509 (±0.125)
Ours	17.8144 (±3.635)	0.6725 (±0.131)

Table 4.1: PSNR and SSIM reported on synthesized visible images obtained using our proposed approach as well as the image synthesis baselines.

4.6 Cross-spectrum face recognition evaluation

The main motivation of the work presented in this chapter is to provide an efficient and prompt solution to integrate thermal technology in already deployed face recognition systems. In this section, we evaluate the efficiency of our proposed approach of thermal-to-visible image synthesis in context of cross-spectrum face recognition. Firstly, we introduce the algorithms selected to carry out face recognition experiments. Then, we define the experimental scenarios that we have considered. Finally, results and discussion are presented.

4.6.1 Face recognition algorithms

For evaluating the synthesized faces when used in cross-spectrum face recognition task, we measured the recognition performance of two selected widely-used face recognition algorithms:

OpenFace [109] is an implementation of face recognition system using deep neural networks based on Google's FaceNet [110] architecture. The OpenFace network is trained using the combination of the two largest public face databases CASIA-WebFace [111] and FaceScrub [112]. The training of the OpenFace model was based on triplet loss minimization. The evaluation of OpenFace model provided competitive performances compared to previous state-of-the-art systems. We use the OpenFace pretrained model to map faces into 128-dimension embeddings. Then, nearest neighbours algorithm is applied using the Euclidean distance to discriminate matching samples.

LightCNN [113] is a new implementation of CNN for face recognition designed to have fewer trainable parameters and to handle noisy labels. This network introduces a new concept of max-out activation in each convolutional layer, called Max-Feature-Map, for feature filter selection. This network has achieved better performance than CNNs while reducing computational costs and storage space. When evaluated on the LFW database, LightCNN achieved face recognition accuracy of 99.33%, outperforming OpenFace that obtained a 92.92% of accuracy. We used the trained network with 29-layers to obtain embeddings of 256-dimension from face images. Embeddings extracted from gallery and probe templates are compared using cosine similarity.

4.6.2 Experimental scenarios

The performance of our image synthesis solution in cross-spectrum face recognition is compared to face recognition experiments performed in the following scenarios:

Visible: We perform face recognition in the visible spectrum, by considering the neutral face image as gallery and the rest of the facial variations as probe images. This will report the performance of the selected face recognition algorithms, that will be considered as an upper bound for the evaluation of the synthesized images. Besides, this baseline will depict the utility of thermal-to-visible face synthesis in hands-on scenarios, in particular when the face is acquired in poorly lit environments.

Thermal: Here, we conduct cross-spectrum face recognition without any modifications applied to the thermal data. Simply put, we consider as gallery set the neutral face image acquired in visible spectrum and as probe set all the other face variations acquired in thermal spectrum. This baseline will quantify the gap between the two spectra.

Isola et al. [96] (Pix2Pix) We perform cross-spectrum face recognition by matching the synthesized visible faces obtained by the model proposed by Isola et al. [96] against the visible face enrollments. It is interesting to compare our approach to this baseline, as it is considered a benchmark for image synthesis.

Zhang et al. [84] (TV-GAN) Synthesized face images obtained by the model proposed by Zhang et al. [84] are matched against visible face enrollments. The performance reported by the face recognition algorithms when using the synthesized face images obtained by TV-GAN will quantify the improvement brought by appending the identity loss term in the training of the Pix2Pix model. In addition, evaluating the model proposed

by Zhang et al. [84] will lead to a fair comparison of our approach, as both models are introduced in the same framework, i.e. that of thermal-to-visible image synthesis.

4.6.3 Experimental setup

The database contains in total 21 different facial variations. Cross-spectrum face recognition evaluation is performed for the different variation set separately. Therefore, we have split the database into 5 subsets of as follow:

Neutral	1 sample/subject
Expression	6 samples/subject
Head pose	4 samples/subject
Occlusion	5 samples/subject
Illumination	5 samples/subject

Table 4.2: Distribution of the database across the defined subsets.

The neutral face image acquired in visible spectrum is considered as an enrollment sample for all the subjects.

4.6.4 Results

In order to evaluate the synthesized visible face images, we have performed cross-spectrum face recognition using two different systems. The evaluation experiment consists in comparing, in the first place, the synthesized neutral face against the ground truth, and then matching the synthesized faces from each of the facial variation subsets against the visible neutral face. We report, in Table 4.3 and Table 4.4, the recognition accuracy of the OpenFace and LightCNN, respectively. To get a deeper understanding of the performance of the two face recongition systems used to evaluate the results obtained, we plot the receiver operating characteristic (ROC) curves, in Figure 4.7 and Figure 4.8, corresponding to some selected samples from different face variations. It is worth noting that the LightCNN face recognition system results outperform by far that of OpenFace.

We note from the reported results that all synthesis models outperformed the system defined in the thermal scenario, which proves the efficiency of synthesizing visible face images in reducing the spectral gap between visible and thermal data. TV-GAN reports better performances than Pix2Pix confirming the efficacy of the identity loss in preserving the subject identity when synthesizing visible images. Foremost, our proposed solution, based on CRNs, outperforms all the models by a large margin, particularly observed

	Visible	Thermal	Isola et al. [96]	Zhang et al. [84]	Ours
Neutral	-	4	8	20	**20**
Expression	97.66	3.33	7.66	11	**17.33**
Head Pose	75.5	2.5	4	8	**9.5**
Occlusion	80	2	7.2	8.4	**10**
Illumination	80.8	3.2	10.4	11.6	**20**
Average	86.79	3.01	8.49	10.76	**15.37**

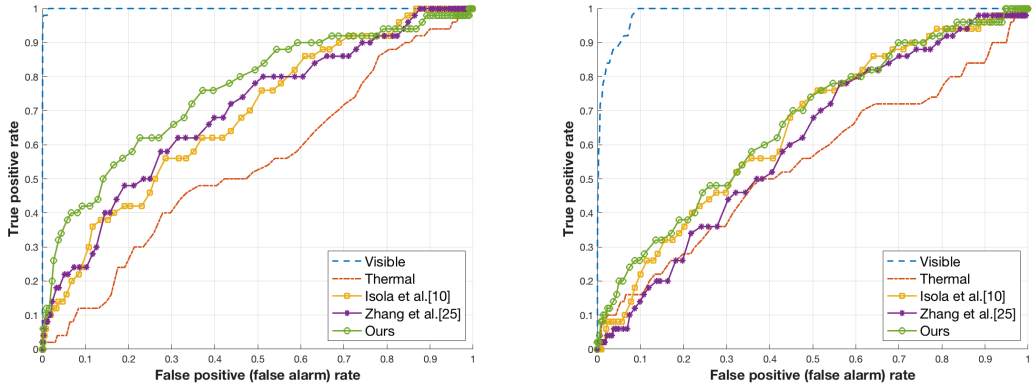
Table 4.3: Cross-spectrum face recognition accuracy across multiple facial variations using OpenFace system

	Visible	Thermal	Isola et al. [96]	Zhang et al. [84]	Ours
Neutral	-	32	48	54	**82**
Expression	99.66	23	37.33	38.33	**67.66**
Head Pose	80.5	12.5	14.5	15.5	**30**
Occlusion	98.8	14.4	16.4	25	**44.8**
Illumination	87.2	15.6	29.6	35.2	**63.6**
Average	95.232	19.5	29.166	33.606	**57.612**

Table 4.4: Cross-spectrum face recognition accuracy across multiple facial variations using LightCNN system

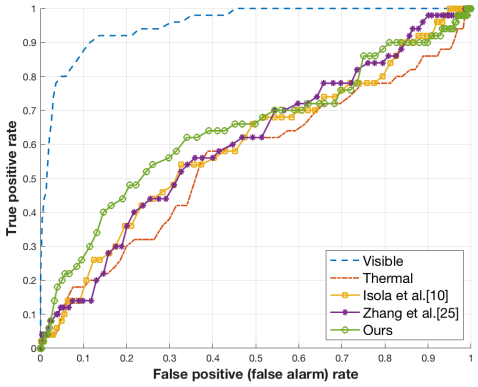
on LightCNN results, and that applies to all facial variations. This is mainly due to the limitations of GANs that are known for being data hungry. However, our system succeeded in generating relatively high quality visible images despite the limited size of the training data. Furthermore, both Pix2Pix and TV-GAN models are trained using a L1 loss function, making them very sensitive to image misalignment. Alternatively, our proposed system uses contextual loss which makes it inherently scale and rotation invariant.

The improvements in performance reported by our proposed approach, is relatively higher for neutral, expression and illumination variations when compared to the improvements in performance reported on occlusion and head pose variations. This is due to the fact that our proposed model of thermal-to-visible face synthesis, as well as the two baseline models, are more likely to fail in generating correct facial traits when the face is presented in a challenging head pose and/or occlusion variations.



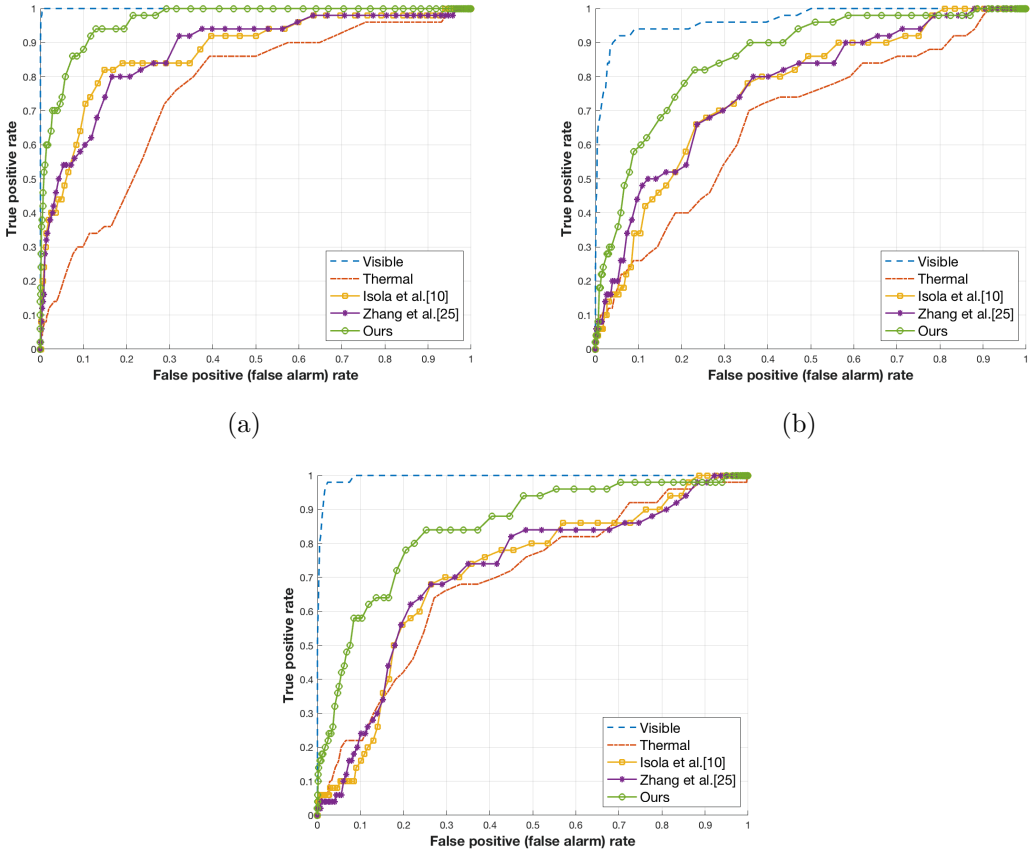
(a)

(b)



(c)

Figure 4.7: ROC curves of cross-spectrum face recognition based on OpenFace system for selected samples from: (a) expression variation, (b) head pose variation, (c) occlusion variation.



(a) (b)

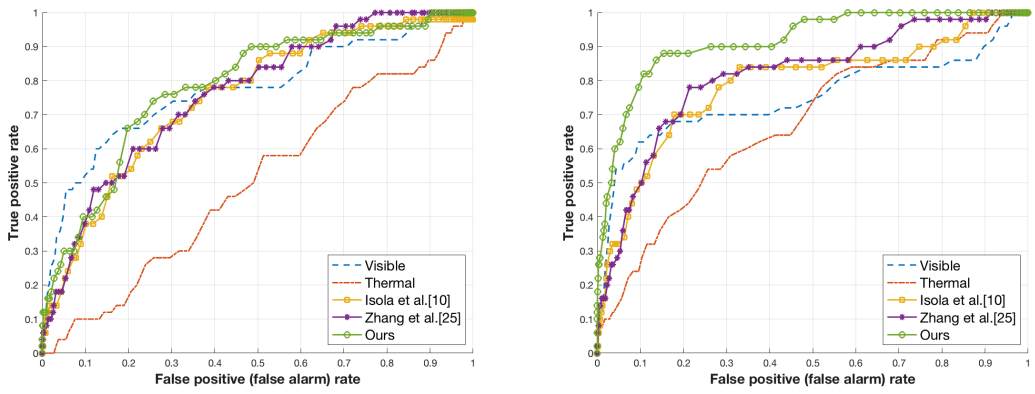
(c)

Figure 4.8: ROC curves of cross-spectrum face recognition based on LightCNN system for selected samples from: (a) expression variation, (b) head pose variation, (c) occlusion variation.

Table 4.5 reports the rank-1 recpgnition of OpenFace and LightCNN face recognition systems when employed in total darkness. We plot also, in Figure 4.9, the ROC curves of the two evaluation systems in the absolute dark condition. We can clearly observe that our proposed model not only outperforms other face synthesis models but also it provides significantly higher performance compared to the visible spectrum. This affirms the efficacy of face synthesis from thermal to visible in one of the most challenging scenarios such as poorly lit environments.

	Visible	**Thermal**	**Isola et al. [96]**	**Zhang et al. [84]**	**Ours**
OpenFace	16	2	10	14	**22**
LightCNN	42	16	28	36	**56**

Table 4.5: Cross-spectrum face recognition accuracy in operative scenario where samples were acquired in total darkness.



(a) (b)

Figure 4.9: ROC curves of cross-spectrum face recognition in dark environment: (a) OpenFace system (b) LightCNN system.

4.7 Summary

Although several efforts have been devoted in recent years for face synthesis from thermal to visible spectrum, the task remains challenging considering the shortage of the available data designed for this task. We present, in this chapter, a novel solution based on cascaded refinement networks, that succeeded in generating color visible image of satisfying quality, trained on a limited size database. The proposed network is based on the use of a contextual loss function, enabling it to be inherently scale and rotation invariant. Despite the existence of challenging facial variations such as occlusions, expression, head pose and illumination, our solution has produced the most visually pleasing synthesized face

images when compared to existing work. We also performed an evaluation of our solution in cross-spectrum face recognition task. The reported results have shown that our system outperforms recent face synthesis systems. Underlining the motivation of face synthesis from thermal to visible spectrum, we have proved that face recognition performance reported on the synthesized images is significantly higher than the one reported on visible spectrum when operated in poorly lit environments, as it was improved by 37.5% (i.e. from 16% to 22%) and 33.33% (i.e. from 42% to 56%) evaluated by OpenFace and LightCNN, respectively.

Chapter 5

Illumination invariant face recognition based on dynamic quality-weighted fusion of visible and thermal spectrum

A new scheme of score level fusion is introduced in this chapter for illumination invariant face recognition from visible and thermal spectrum. The work presented in this chapter explores a direction leading to a fast and smooth integration into existing face recognition systems and does not require recollection of enrollment data in thermal spectrum. This chapter investigates the potential role of thermal spectrum in improving face recognition performances when employed under adversarial acquisition conditions. We consider a context where individuals have been enrolled solely in visible spectrum, and their identity will be verified using two sets of probes: visible images and thermal-to-visible images. The thermal-to-visible face synthesis [114] is performed using the approach presented in Chapter 4, and face features are extracted and matched using LightCNN [113] and Local Binary Patterns [115]. The contribution of this work lies in performing the fusion procedure through several quality measures computed on both visible and thermal-to-visible synthesized probes and compared to the quality of visible gallery images, in a way that it determines the relevance of each of the probes in improving the face recognition performance

The remainder of this chapter is organized as follows. Motivations leading to this work are given in section 5.1. A literature overview on visible and thermal spectrum fusion, followed by a brief review on quality based fusion for multimodal biometric, is presented in

Section 5.2. Section 5.3 introduces the proposed dynamic quality-based fusion scheme for illumination invariant face recognition. Experimental results are presented in Section 5.4. A summary of the work and findings reported in this chapter ares given in Section 5.5.

5.1 Motivation

Our first attempt to synthesize visible face images from thermal inputs [114] took the first steps towards enabling a prompt and easy integration of thermal sensors in already deployed face biometric systems. While this work showed improvement in performance in terms of visual quality [116] and cross-spectrum face recognition compared to some selected baseline models [84, 96], face recognition based solely on visible spectrum significantly outperforms systems based on synthesized visible face images when operated under controlled illumination conditions.

It is undeniable that face recognition performance reported on the synthesized images is significantly higher than the one reported on visible spectrum when engaged in poorly lit environments, as face recognition accuracy was improved by 37.5% [114] for LightCNN system. However, synthesized visible face images are stillfew steps behind compared to visible images when confronting other sorts of variations. Thermal-to-visible face synthesis inadvertently generates few artctifacts and occasionally some wrong facial attributes that may alter the face matching process. In an attempt to achieve an illumination invariant face recognition system operating continuously day and night, we propose to fuse scores obtained while matching visible face probes with visible face gallery and the scores obtained by matching thermal-to-visible generated images from thermal face against the same visible face gallery. Based on the intuition that image quality can be indicative of the utility of a face sample, we propose to fuse the score of matching visible face images and synthesized visible face images against visible gallery images, based on the image quality score of each component.

5.2 Related work

Since the emergence of thermal imagery in biometrics, a lot of efforts have been devoted to performing visible and thermal fusion in order to achieve improvements in unconstrained face recognition research. Several studies [46, 47] explored the usage of genetic algorithms (GAs) to select features extracted separately from visible and thermal spectra and perform fusion at score level. Desa et al. [117] used GAs to find the optimal strategy of feature fusion at non linear transformed domain, exploring two non linear face subspaces: Kernel Principle Component Analysis and Kernel Fisher's Linear Discriminant Analysis. Chen et

al. [118] used a decision based fuzzy integral fusion of visible and thermal face recognition results. Buyssens et al. [119] introduced a special type of CNN based on diabolo network model [120] to extract features from both visible and thermal images and then fused the matching scores. Hariharan et al. [39] proposed a new data-level fusion scheme using empirical mode decomposition.

In an attempt to exploit the thermal spectrum for illumination invariant face recognition, several fusion studies have been proposed. Heo et al. [48] proved the complementarity of visible and thermal spectrum for illumination invariance by investigation data and decision level fusion. Arandjelovic et al. [33, 34, 121] presented a multistep fusion scheme, carried out at the decision level and holistic and local feature level of visible and thermal faces. Socolinsky et al. [30, 122, 123] proposed a simple decision based fusion using a weighted combination of visible and thermal matching scores. The proposed fusion scheme was evaluated indoors and outdoors, resulting in better face recognition performance in varying illumination conditions but failing in extreme illumination conditions.

The research objective of this work is to provide a continuous face recognition system that is invariant to illumination changes. This can be achieved by setting up a visible and thermal fusion scheme where the weight of each component is assigned by the corresponding image quality.

Several fusion and modality selection solutions were proposed, in setting multimodal biometric systems, based on quality assessment of the biometric sample. Good quality image usually yields a robust matching performance. Fierrez-Aguilar et al. [124] introduced one of the earliest works of biometric quality fusion at the score level, integrating quality information into a Bayesian statistical model for multimodal biometric classification. Using a unimodal biometric system, Vatsa et al. [125] proposed fusing the RGB channels based on quality scores to improve the performance of iris recognition. Zhou et al. [126] presented quality based eye recognition by segmenting the eye into iris ans scelra and performing classification on the selected region as reported by its quality.

5.3 Quality-weighted score fusion

In this section, we describe in detail the proposed fusion solution. First, we depict the continuous day and night face recognition scenario. Then, we define the two face recognition systems used to compare face samples and obtain their matching scores. Subsequently, we list the quality assessment metrics considered in this study. Finally, we describe the proposed quality-weighted fusion scheme.

5.3.1 Scenario description

The main motivation of this work is to assure a continuous day and night face recognition while granting an easy integration of thermal sensors in face recognition systems. The thermal sensor integration is provided by synthesizing visible face images from thermal inputs and matching the synthesized image against the visible gallery samples [114], as presented in Chapter 4. As for the continuity of face recognition, it is controlled by the quality weighted fusion of matching visible faces and synthesized visible faces against visible face gallery. Thereby, the participation of each component is indicated by the corresponding quality score.

Figure 5.1 depicts different gallery samples as well as probe samples in three different illumination conditions. Probe VIS corresponds to face images acquired in visible light spectrum, whereas Probe G_{VIS} represents synthesized visible faces from thermal inputs. Thermal-to-visible face synthesis model [114] is presented in Chapter 4. Training the thermal-to-visible face synthesis model was carried out using numerous facial variations taken in controlled illumination conditions. This model provides a faithful estimation of the visible information based on the thermal input when it is initially missing in the visible spectrum. In other words, this step is essential to provide the missing visible information due to lack of illumination. In case of Condition 1, when the illumination conditions are controlled, the quality of visible images is undoubtedly superior to the quality of synthesized visible images. Consequently, it is expected that the proposed quality based fusion scheme will leverage the visible spectrum to obtain accurate face recognition results. While in Condition 2, some information in visible face images are missing due to low illumination. In this case, our proposed solution is supposed to exploit the information provided by the visible images and the synthesized visible images complementary. In case of Condition 3 however, the visible information is almost completely absent, which may encourage our proposed fusion system to consider for the most part the information obtained from the synthesized visible faces.

5.3.2 Face feature extraction and matching

We present here the face comparison systems used to obtain the matching scores on which the fusion will be applied. We selected a state-of-the-art system based on deep learning embeddings and a second system based on handcrafted features.

LightCNN [113] is a pretrained model of a light CNN of 29 layers. LightCNN was used in Chapter 4 and led to better face recognition performances compared to a similar baseline based on OpenFace [109], and thus it was retained for the work presented in

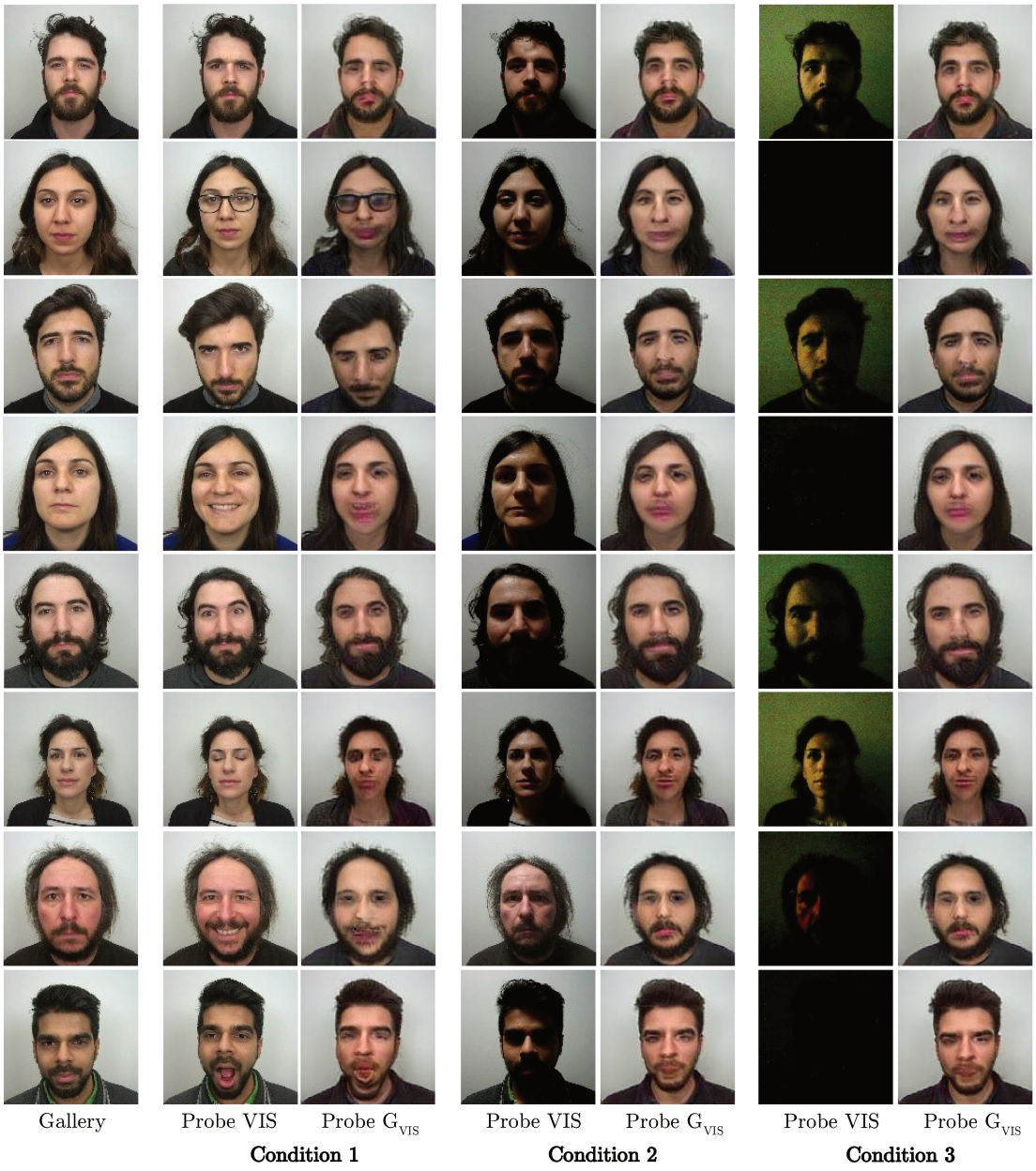


Figure 5.1: Illustration of continuous day and night face recognition scenario under 3 different illumination conditions. Condition 1: controlled illumination environment, condition 2: low illumination environment, condition 3: extremely poor illumination environment.

this chapter as well. 256-dimension embeddings are extracted, using LightCNN, from gallery and probe samples and then matched using cosine similarity.

Local Binary Pattern (LBP) was originally introduced by Ojala et al. [115] for texture analysis, but later on it was thoroughly explored in numerous applications. Particularly, it has shown its efficiency for face analysis not only in visible but also in thermal spectrum. LBP represent a binary pattern that describes the local neighborhood of each pixel of the face image. The obtained LBP features are then concatenated to create a single histogram feature vector of 256-dimensions. Histograms extracted from gallery and probe image samples are compared using the χ^2 distance as dissimilarity measure.

5.3.3 Quality assessment metrics

Most often, quality of face samples reflects their relevance in providing a correct and accurate recognition with a high matching score. High quality samples often deliver highly informative features, yet low quality samples suffer heavily from noisy data and missing information. Therefore, selecting quality assessment metrics is very critical in boosting or lowering recognition performance.

We present, here, a number of selected quality metrics in order to study the impact of each on face recognition performance.

- **Lightening symmetry [127]:** it quantifies the symmetry between sub-regions of an image and can be measured as the difference between the histogram of intensity in each half sub-region.
- **Brightness [128]:** is given by the average value of the image intensity histogram.
- **Contrast [128, 129]:** can be defined as the scale difference between maximum and minimum intensity values in an image.
- **Global Contrast Factor (GCF) [130]:** is the weighted sum of local contrast for various resolutions of the image.
- **Exposure [131]:** indicates the amount of light in the image and can be measured using image statistical measures.
- **Blur [132]:** is based on the fact that sharp images have thin edges and blurry images have wider edges, blur is expressed as the edge width.
- **Sharpness [129]:** is defined as the sum of gradients at every pixel intensity.

5.3.4 Proposed fusion scheme

Figure 5.2 illustrates the proposed asymmetric approach of quality weighted fusion at score level. Let Q_{VIS}, $Q_{G_{VIS}}$ and $Q_{Gallery}$ denote the quality measures of the visible image probe, the quality of the thermal-to-visible generated image probe, and the quality of visible gallery image, respectively, obtained using one of the quality assessment metrics just presented in Section 5.3.3. During recognition, we calculate the quality similarity scores of the original visible image and the thermal-to-visible synthesized image by determining their similarity to $Q_{Gallery}$, as follow:

$$QS_i = e^{\frac{Q_{Gallery} - Q_i}{Q_{Gallery}}}, \, where \, i \in \{VIS, G_{VIS}\}. \tag{5.1}$$

Once the quality scores are obtained, they are normalized using min-max normalization. Then, we compute the weight to be assigned to each entity, as

$$w_i = \frac{QS_i}{QS_{VIS} + QS_{G_{VIS}}}, \; i \in \{VIS, G_{VIS}\}. \tag{5.2}$$

The closer Q_i is to $Q_{Gallery}$, the higher the weight will be assigned to i. Next, the face matching scores, denoted by S_i, are computed. S_{VIS} are obtained by comparing the visible image probe to the visible gallery set. $S_{G_{VIS}}$ are calculated by performing a face comparison between the synthesized visible image and the visible gallery set. The obtained matching scores are then normalized. The overall fused score is computed using the weighted exponential sum rule, as follow:

$$S_{fused} = \sum_i w_i e^{S_i}, \, where \, i \in \{VIS, G_{VIS}\}. \tag{5.3}$$

Simply put, the quality weight will play a role in determining whether the visible sample is reliable enough to provide an accurate recognition. The quality of visible samples deteriorates mostly due to lack of illumination. Thereupon, the proposed fusion scheme will favor the synthesized visible sample as it is estimated from thermal inputs that are immune to illumination variations. The proposed method is summarized in Algorithm 1.

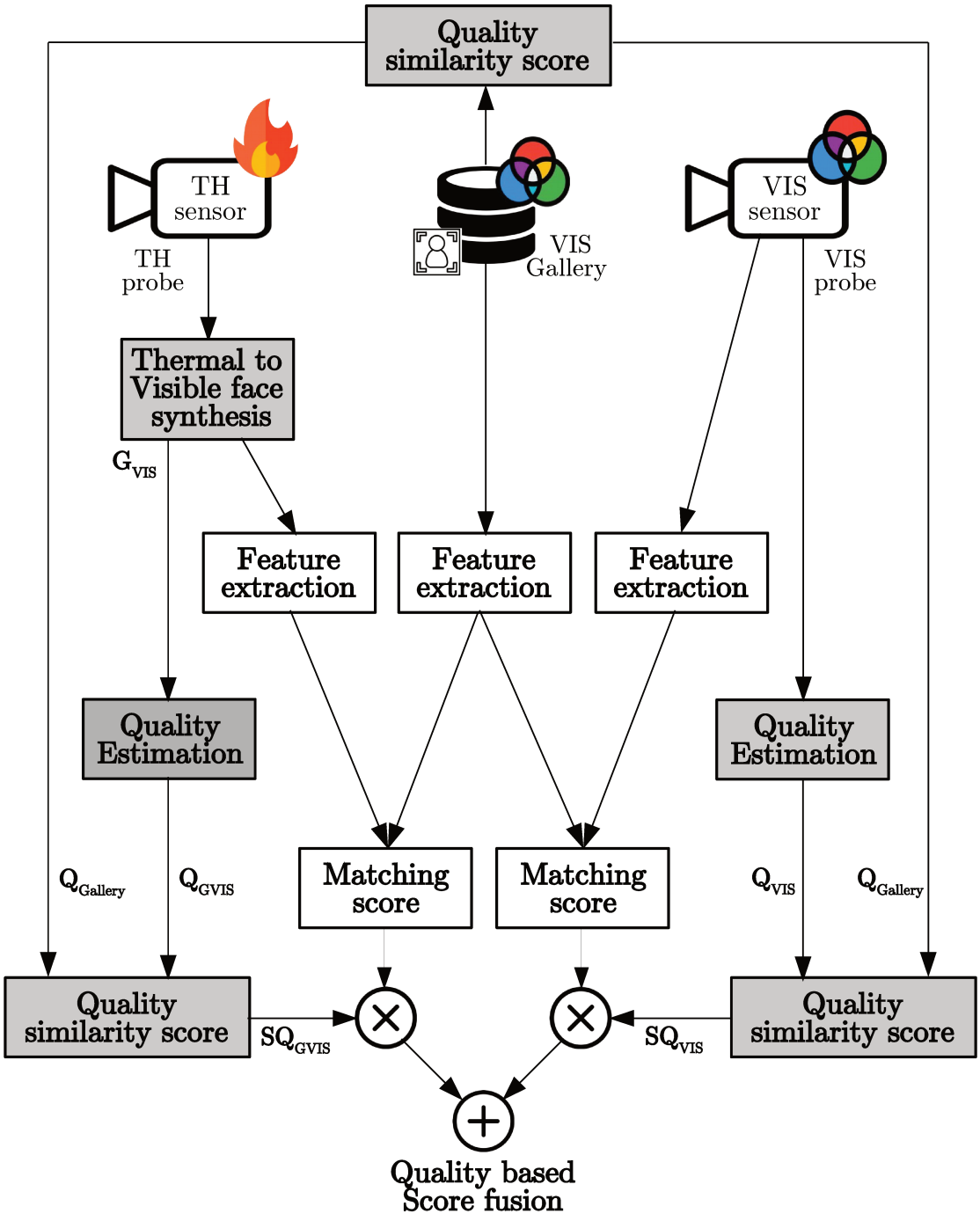


Figure 5.2: Framework of the proposed quality-based score fusion scheme, where *VIS*, *TH* and G_{VIS} denote the visible image, the thermal image and the synthesized visible image from the thermal capture, respectively.

Algorithm 1: Quality-weighted score fusion

Input *Probe Samples:* set of samples acquired simultaneously in visible and thermal spectrum under various facial variations.

Gallery Samples: set of neutral face samples acquired solely in visible spectrum.

for $p \in$ *Probe Samples* **do**

- $VIS \leftarrow$ Read Visible Image (p)
- $TH \leftarrow$ Read Thermal Image (p)
- $G_{VIS} \leftarrow$ Thermal-to-Visible face synthesis (TH) as per chapter 4.
- $Q_{VIS} \leftarrow$ Quality Estimation(VIS)
- $Q_{G_{VIS}} \leftarrow$ Quality Estimation (G_{VIS})
- **for** $g \in$ *Gallery Samples* **do**
  - $Gallery \leftarrow$ Read Visible Image (g)
  - $Q_{Gallery} \leftarrow$ Quality Estimation$(Gallery)$
  - $QS_{VIS}(p, g) \leftarrow$ Quality Similarity Score $(Q_{VIS}, Q_{Gallery})$ as per Eq.5.1
  - $QS_{G_{VIS}}(p, g) \leftarrow$Quality Similarity Score $(Q_{G_{VIS}}, Q_{Gallery})$ as per Eq.5.1
  - $S_{VIS}(p, g) \leftarrow$ Matching Score $(VIS, Gallery)$ as per Sec.5.3.2
  - $S_{G_{VIS}}(p, g) \leftarrow$ Matching Score$(G_{VIS}, Gallery)$ as per Sec.5.3.2
- **end**

end

Min-Max normalization of QS_{VIS}, $QS_{G_{VIS}}$, S_{VIS} and $S_{G_{VIS}}$

Compute weights w_{VIS} and $w_{G_{VIS}}$ as per Eq.5.2

$S_{fused} \leftarrow$ Quality-weighted score fusion (w_{VIS}, S_{VIS} , $w_{G_{VIS}}$ $S_{G_{VIS}}$) as per Eq.5.3

return *the overall fused score* S_{fused}

5.4 Experiments and results

In this section, we present the data used to perform face recognition based on quality weighted fusion. Then, we detail the evaluation protocol used to assess the proposed fusion approach. Finally, we present the obtained results followed by an analysis of the impact of different quality assessment metrics on face recognition performance.

5.4.1 Database

We used the VIS-TH face database [105], presented in Chapter 3, for the evaluation of our proposed fusion solution. Three different sets are considered:

- **Gallery set:** face samples acquired in visible spectrum under controlled illumination conditions, with neutral expression and frontal head pose.
- **Probe VIS:** probe face samples acquired in visible spectrum under different facial variations including varying illumination conditions.
- **Probe G_{VIS}:** probe face samples initially acquired in thermal spectrum under different facial variations including varying illumination conditions, and then converted into visible spectrum. Thermal-to-visible face synthesis is detailed in section 4.3 of chapter 4.

5.4.2 Experimental protocol and results

Feature extraction is performed using either LBP or LightCNN. Feature vectors from gallery and probe sets are compared to obtain the matching scores of the two components. In parallel, quality measures are computed using 7 different quality assessment metrics and quality similarity scores are then deduced. Dynamic quality weighted fusion at score level is carried out as described in Section 5.3.4. The performance of our proposed fusion approach is compared to the performance of fusing scores obtained from matching visible probes and thermal probes against a common visible gallery set.

To highlight the main motivation of thermal spectrum usage in face recognition, we display, in Figure 5.3, the receiver operating characteristic (ROC) curve of the three setups aforementioned for face images that were acquired in total darkness. We can clearly observe that the setup based on thermal-to-visible synthesized images provides significantly higher performance compared to the setup based on visible images. This affirms the efficacy of thermal imagery in most of the challenging scenarios, i.e. poorly lit environments. Also, we note that the setup based on thermal-to-visible synthesized images

outperforms the thermal based setup, which proves the efficiency of thermal-to-visible face synthesis in reducing spectral gap between visible and thermal spectrum.

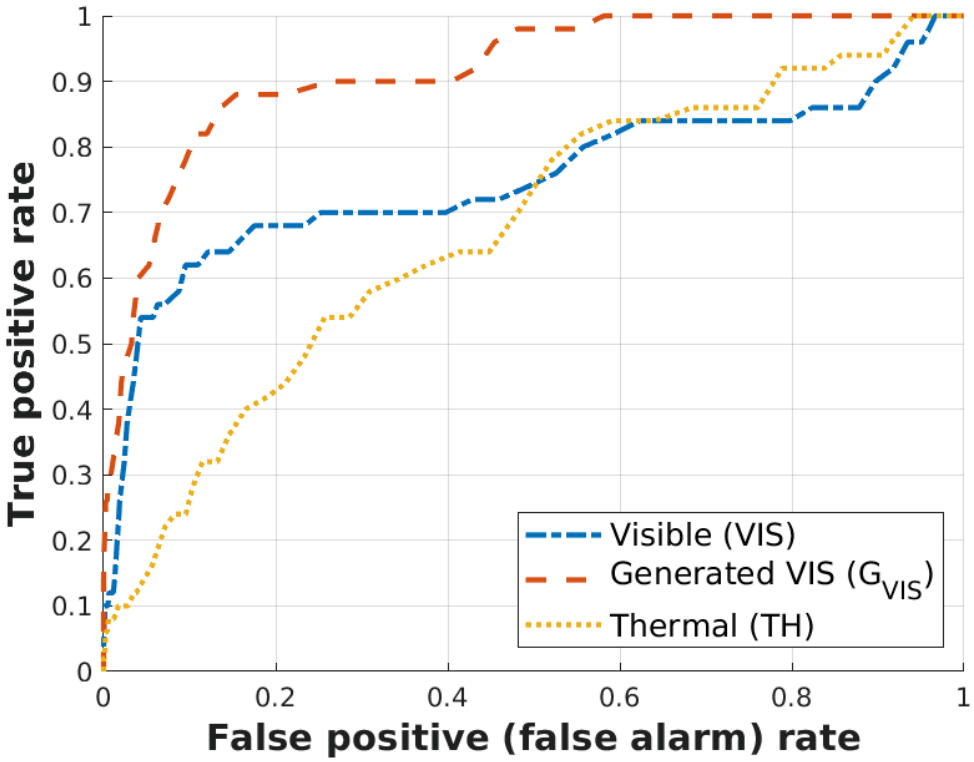


Figure 5.3: ROC curves in extremely poor illumination environment using LightCNN system

Table 5.1 presents the rank-1 recognition of LightCNN and LBP systems reported over all the facial variations contained in VIS-TH database. In this table, we report firstly the recognition performance of each of the following setups: matching visible probe, original thermal probe and thermal-to-visible synthesized faces against visible gallery. We observe that face recognition using the synthesized visible images leads to better performance than when using thermal images, which proves the efficiency of thermal-to-visible face synthesis in reducing the gap between visible and thermal spectra. Although, the synthesized visible images are still few steps behind standard visible face

images and that is perceivable mostly for the performance across all the facial variations.

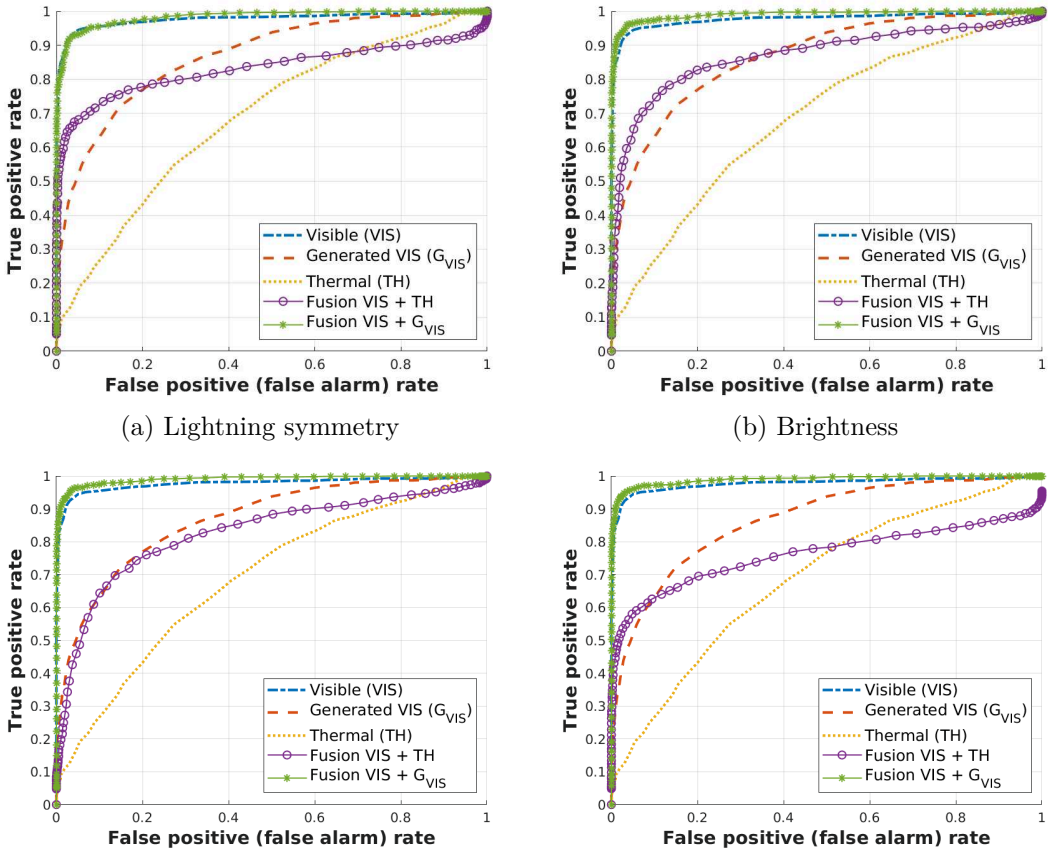
Furthermore, we can evidently perceive that face recognition using deep learning embeddings (LightCNN) outperforms hand-crafted features (LBP) which confirms the assertions presented in [133].

To assess the impact of each quality metric used in this Chapter, we report rank-1 recognition of quality weighted fusion of visible images and synthesized visible images (denoted as ***(VIS, G_{VIS})*** in Table 5.1) for each quality metric, where Q^1, Q^2, Q^3, Q^4, Q^5, Q^6 and Q^7 denote lightning symmetry, brightness, contrast, GCF, exposure, blur and sharpness, respectively. Q^{avg} refers to using the average quality score of the 7 quality assessment metrics. Furthermore, quality weighted score fusion of visible face images and original thermal images (denoted as ***(VIS,TH)*** in Table 5.1) is considered as a baseline. We note that the described fusion scheme using the thermal-to-visible face synthesis unit outperforms significantly the fusion of visible and thermal images plainly. This divergence in performance certifies the proficiency of thermal-to-visible face synthesis in bringing the two spectra closer together. The rank-1 recognition results of LightCNN system showed that the proposed fusion approach has led to the best performance, particularly for global contrast factor quality metric. However, we can determine that the proposed quality weighted score fusion shows nearly similar performance for all the quality assessment metrics.

To get a deeper understanding of the performance of our proposed fusion scheme, we plot the ROC curves, in Figures 5.4 and 5.5. The ROC curve is computed over all the facial variations contained in the database, so as to demonstrate the efficacy of our proposed approach in a wide range of operative scenarios. The plot confirms our previous observations, as we can see that all the considered quality assessment metrics impact the performance of the fused system similarly. Conclusively, we observe that the proposed fusion based approach in this chapter outperforms face recognition operating solely on visible data. It is fair to admit that the difference of performance is not significantly large, that is due to the distribution of the variations within the database, as it contains more samples acquired under controlled illumination conditions compared to only few samples acquired under low illumination conditions that highlights the thermal imagery usage.

	LightCNN					LBP				
	VIS	**TH**	**G_{VIS}**	**(VIS, TH)**	**(VIS, G_{VIS})**	**VIS**	**TH**	**G_{VIS}**	**(VIS, TH)**	**(VIS, G_{VIS})**
$\mathbf{Q^1}$	0.916	0.180	0.542	0.643	0.880	0.821	0.042	0.457	0.211	0.638
$\mathbf{Q^2}$				0.805	0.921				0.284	0.698
$\mathbf{Q^3}$				0.775	0.923				0.440	0.729
$\mathbf{Q^4}$				0.508	0.925				0.363	0.696
$\mathbf{Q^5}$				0.542	0.918				0.337	0.718
$\mathbf{Q^6}$				0.805	0.92				0.28	0.702
$\mathbf{Q^7}$				0.735	0.908				0.428	0.680
$\mathbf{Q^{avg}}$				0.746	0.923				0.429	0.735

Table 5.1: Rank-1 recognition across multiple facial variations using LightCNN and LBP face recognition algorithm.



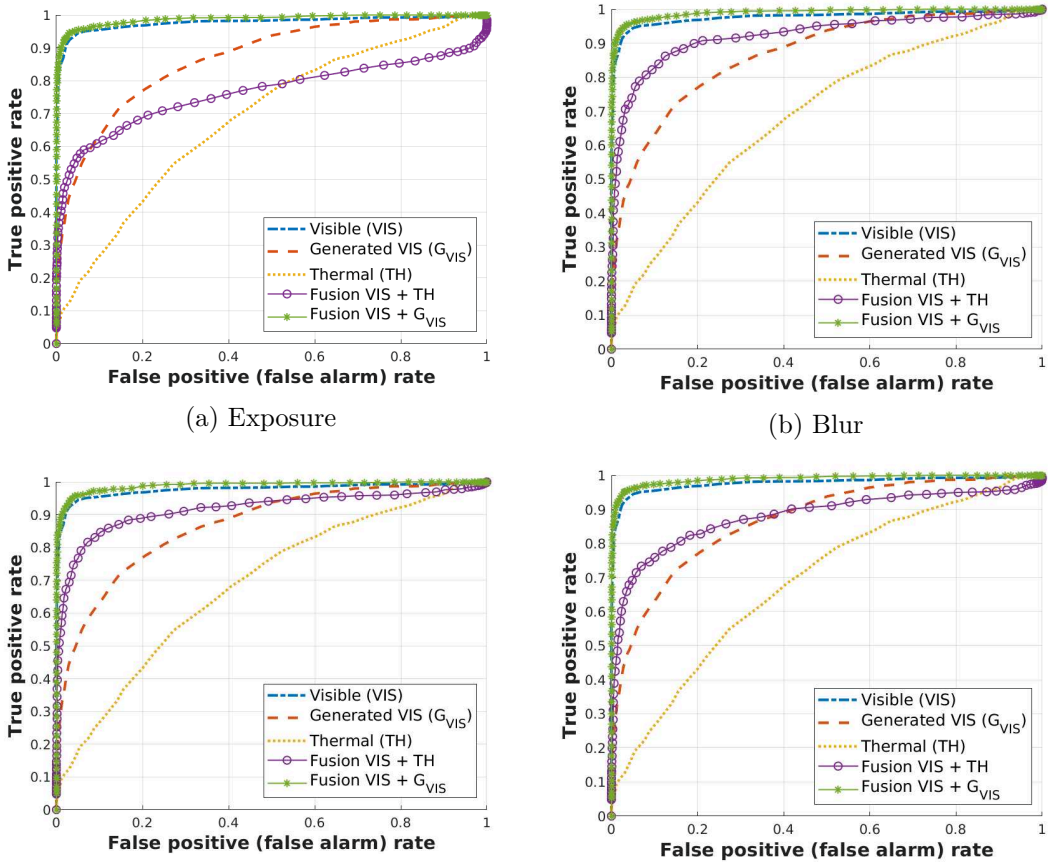
(a) Lightning symmetry

(b) Brightness

(c) Contrast

(d) Global Contrast Factor (GCF)

Figure 5.4: ROC curve deduced over all the facial variations in VIS-TH database [105] using LightCNN



(a) Exposure

(b) Blur

(c) Sharpness

(d) Average

Figure 5.5: ROC curve deduced over all the facial variations in VIS-TH database [105] using LightCNN

.

5.5 Summary

Integrating thermal imagery in face recognition systems tackles, particularly, the poor illumination challenge for visible spectrum. Therefore, a new scheme of score level fusion for robust face recognition from visible and thermal face data that enables straightforward integration in the existing face recognition systems is proposed in this chapter. The proposed system operates according to the following protocol in face recognition: individuals have been enrolled solely in visible spectrum (i.e. gallery) but can be afterwards controlled by dual visible and thermal acquisition (i.e. probe). Considering that the gap between the visible and thermal spectra is important, it was necessary to include a step where synthesized visible images are generated from thermal inputs. This solution benefits from the quality measures of the visible gallery and probe faces to assign weights for visible and thermal samples in order to provide an illumination invariant face recognition solution. The results report an interesting improvement in face recognition performance compared to when using solely visible samples. In addition, results have proved the efficiency of thermal-to-visible face synthesis in providing more accurate performance for face recognition system.

Chapter 6

Facial landmark detection on thermal data through fully annotated thermal data synthesis

Facial landmark detection is a crucial prerequisite for facial image processing. Given the upswing of deep learning based approaches, the performance of facial landmark detection has been significantly improved. However, this uprise is mostly limited to visible spectrum based face analysis tasks, as there are only few research works on facial landmark detection in thermal spectrum. This limitation is mainly due to the lack of available thermal face databases that include full facial landmark annotations. In this chapter, we propose to tackle this data shortage by converting existing face databases, designed for the facial landmark detection task, from visible to thermal spectrum. By doing so, facial landmark annotations available in databases collected in the visible spectrum can be leveraged in their artificially generated, thermal, counterpart. Using the synthesized thermal databases along with the facial landmark annotations, two different facial landmark detection models are trained using active appearance models [134] and deep alignment networks [135]. The evaluation of these models shows accurate facial landmark detection on real thermal data of different quality. With the need to provide prompt solutions for thermal face analysis, our proposed framework provides a vehicle to fuel future research in thermal imagery, not only limited to facial landmark detection but also extendable to other tasks that require extensive annotation.

The remainder of this chapter is organised as follows. Section 6.1 introduces the motivation behind this work. Section 6.2 presents the previous work in facial landmark detection mainly focused on thermal spectrum. Section 6.3 describes the selected

databases to synthesize a thermal face database and the employed landmark annotation standard, followed by a presentation of the proposed approach to perform visible-to-thermal face synthesis. Section 6.4 introduces two selected approaches used for facial landmark detection in this work. Section 6.5 reports the experimental setup and the evaluation protocol followed by results and discussion. A summary is presented in Section 6.6.

6.1 Context and motivation

Facial landmark detection (FLD) consists in locating predefined landmarks, such as eye contours, eye brows, nose, lips in a human face. These detectors provide a shape representation of the face that captures transformations due to facial expressions and/or head movement. FLD has drawn a lot of attention during recent times, as it became an essential requirement to perform a wide range of task related to facial image processing, e.g. face alignment and frontalization [136,137], 3D face reconstruction [136,138], emotion recognition [139] and lip reading [140]. However, FLD on thermal data has not been extensively explored yet, and to our knowledge there are no public facial landmark detectors available designed for thermal spectrum. Thermal imagery provides data with lower spatial resolution and contrast when compared with visible imagery, and it also lacks textural and geometrical information. Therefore, applying the advances of FLD designed for visible data to thermal spectrum may be challenging. Also, the lack of public thermal face databases available with facial landmark annotations prevents thermal spectrum from benefiting from the recent advances in deep learning that have led to remarkable improvements in FLD performance, including when tested *in-the-wild*.

In this work, we present a novel concept that aims to tackle the lack of annotated data in spectra that are less studied than visible spectrum through interspectral conversion, with a focus in the thermal spectrum for FLD task. This proposed concept will enable broader exploration of thermal image processing. Thereby, we provide thermal face databases with full facial landmark annotation through artificial visible-to-thermal data synthesis using existing visible face database designed for FLD, notably LFPW [141] and Helen [142] databases. We explore the possibility of training different FLD models on the synthesized thermal face data to be robust when tested on real thermal data. In particular, we used active appearance models [134] and deep alignment networks [135] to train our facial landmark detectors.

6.2 Related work

FLD in visible spectrum has been extensively studied during the few last decades and it has witnessed great progress. Early works, based on classic parameterized approaches, include active appearance models [134] and constrained local models [143]. Later on, FLD approaches based on cascaded shape regression [144, 145] were introduced. Recently, approaches based on deep learning have achieved impressive results, notably Deep Alignment Network [135] and Style Aggregated Network [146]. A thorough survey of existing techniques of FLD on visible images and videos can be found in [147].

Very few works have focused on FLD on thermal data despite the attention that is being drawn to the usage of thermal imagery in face analysis tasks. First attempts aimed to perform single landmark detection. Tzeng et al. [148] used video frames to detect nostrils through tracking the temperature variation due to respiration. Wang et al. [149] trained a support vector machine (SVM) to perform binary classification of the eye region based on Haar-like features. Alkali et al. [150] located the temperature maxima as it is commonly situated in the inner corner of the eyes.

More recent works focused on the face region as a whole and aimed to detect multiple facial landmark points. Kopaczka et al. [151] trained an active appearance model using histogram of oriented gradients HOG and Scale-invariant feature transform SIFT to perform face tracking in thermal videos. This work has been extended [152] by incorporating the active appearance model into a deep convolutional network to provide it with a prior shape information. These two approaches were trained on a fully annotated thermal face database [153] collected by the University of Aachen. This database provides high spatial resolution data at 1024×768 pixels, with high contrast and noise equivalent temperature difference (NETD) lower than 30mK, meaning that the sensor with which the data is acquired is able to identify very small differences of temperature as little as 30mK or lower. These data specifications result in extremely high quality thermal data much higher than the data provided by the currently available thermal databases and the affordable thermal sensors available on the market. The high quality of the training data of the FLD model mentioned above results in a drastic decrease of landmark detection accuracy when tested on low or medium quality thermal data, which is usually used nowadays for research and commercial purposes.

6.3 Thermal face database synthesis

Several face databases were used in the work presented in this chapter. As a matter of convenience, we gathered all the relevant information about these databases, in Table 6.1, as well as its usage throughout this work.

Database	Spectrum	Thermal spatial resolution	NETD	Facial landmark annotation	Usage
LFPW [141]	Visible	-	-	Provided	Section 6.3.1: Used as input to synthesize thermal data + the provided facial landmark annotation.
Helen [142]	Visible	-	-	Provided	Section 6.3.1: Used as input to synthesize thermal data + the provided facial landmark annotation.
VIS-TH [105]	Visible and thermal	160x120	<100mK	Not provided	Section 6.3.2: Training visible-to-thermal data synthesis.
Aachen database [151]	Thermal	1024 x 768	<30mK	Provided	Section 6.5.1: Training baseline models.
CSMAD	Visible and thermal	320x240	<70mK	Not provided, but possible	Section 6.5.4: Quantitative evaluation on low quality thermal data.
Aachen expression subset [153]	Thermal	1024x768	<30mK	Provided	Section 6.5.5: Quantitative evaluation on high quality thermal data.
UND-X1 [62]	Visible and thermal	320x240	<100mK	Not provided	Section 6.5.6: Qualitative evaluation.
UTW database [154]	Thermal	640x480	<30mK	Not provided	Section 6.5.6: Qualitative evaluation.

Table 6.1: Properties of face databases used in this chapter.

In this section, we describe the selected visible face databases provided with landmark annotation that are used in this work. Then, we describe the approach to perform visible-to-thermal data synthesis in order to obtain a synthesized thermal face database with full facial landmark annotations. Finally, we present some samples of the generated thermal faces.

6.3.1 Face databases with full facial landmark annotation

Numerous visible face databases provided with facial landmark annotation are available [141, 142, 155, 156, 157]. We present, here, the selected databases and the landmark annotation used in this chapter.

Helen [142]: Helen database contains 2330 face images collected from Flickr. The database includes a large set of variations including pose, lighting, expression, occlusion, and individual differences. The facial landmarks were annotated manually using Amazon Mechanical Turk after an initialisation performed using STASM [**?**] algorithm.

LFPW [141]: The Labeled Face Parts in-the-wild database contains 1035 images collected from the web (Flickr, Google, Yahoo...). LFPW database covers the same variations as Helen database. The Labeling and facial landmark annotation were performed by three Amazon Mechanical Turk members.

Facial landmark annotations, used in this work for these databases, were obtained from those released in the context of the *300 Faces in-the-Wild Challenge: the first facial landmark localization Challenge [158].* Organized by *iBUG*[*], the provided annotations attempted to mitigate the mismatched original annotation criterions present in Helen and LFPW databases, with 194 and 29 selected landmark points, respectively. This mismatch in dimensionality motivated the application of a shared semi-supervised approach to FLD followed by manual correction, resulting in a common, consistent, 68 facial points annotation illustrated in Figure 6.1. These annotations, which have been widely used as the *de facto* benchmark for landmark detection, were thus used as reference in the work presented here.

[*]Intelligent Behaviour Understanding Group (iBUG), Department of Computing, Imperial College London

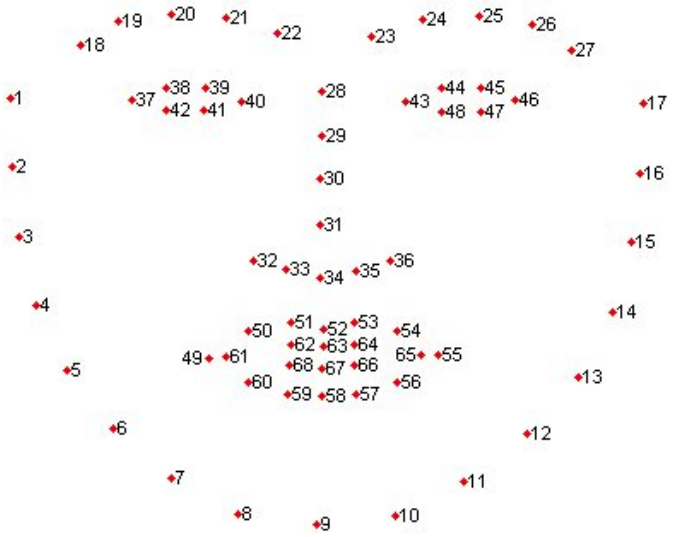


Figure 6.1: 68 facial landmark annotation defined in the context of *300 Faces in-the-Wild Challenge: the first facial landmark localization Challenge [158].*

6.3.2 Visible-to-thermal data synthesis

Data synthesis from visible to thermal spectrum was carried out using the approach presented in Chapter 4. This approach aimed to synthesize visible face images from thermal inputs to perform synthesis based cross-spectrum face recognition. However, in our case, we needed to re-train the model presented in [114] to perform face synthesis in the opposite direction, i.e. from visible to thermal spectrum.

The used approach is based on cascaded refinement networks (CRN) trained using contextual loss, enabling it to be inherently scale and rotation invariant. During the training phase of the visible-to-thermal data synthesis model, we used VIS-Th datatabse [105] introduced in Chapter 3. This database provides thermal images of 160×120 spatial resolution and NETD<100mK acquired with different facial variations. For training, one variation acquired in total darkness was discarded, leaving 1000 pairs of face images. The loss function designed for visible-to-thermal data synthesis is modeled invertedly compared to the loss function defined in Equation 4.5 of Chapter 3. The style loss is computed between the generated thermal image and the ground truth thermal image. Whereas the content loss is computed between the input visible image and the generated thermal image. The training was run for 40 epochs with a learning rate of 1e-4.

To obtain the synthesized databases from visible to thermal spectrum, the images of HELEN and LFPW databases are fed to our trained model, that returns the thermal

version of the input image. Figure 6.2 illustrates some samples of the synthesized thermal face images and their original counterpart. It is worth noting that the synthesized thermal images present realistic pattern of thermal signature. Some details, such as hair, eye brows and teeth, are converted into high pixel values reflecting regions with lower temperature compared to the face region. In addition, the nose region is generated slightly darker, as the nose is usually colder than the rest of the face because it is composed mainly of cartilaginous tissue. Also, eyes contours are generated lighter than the rest of the face, which reflects realistic thermal signature as the high temperatures are usually situated around the eye region. The synthesized images also present some artefacts as we can observe, in few samples, dark patterns located at arbitrary regions of the face.

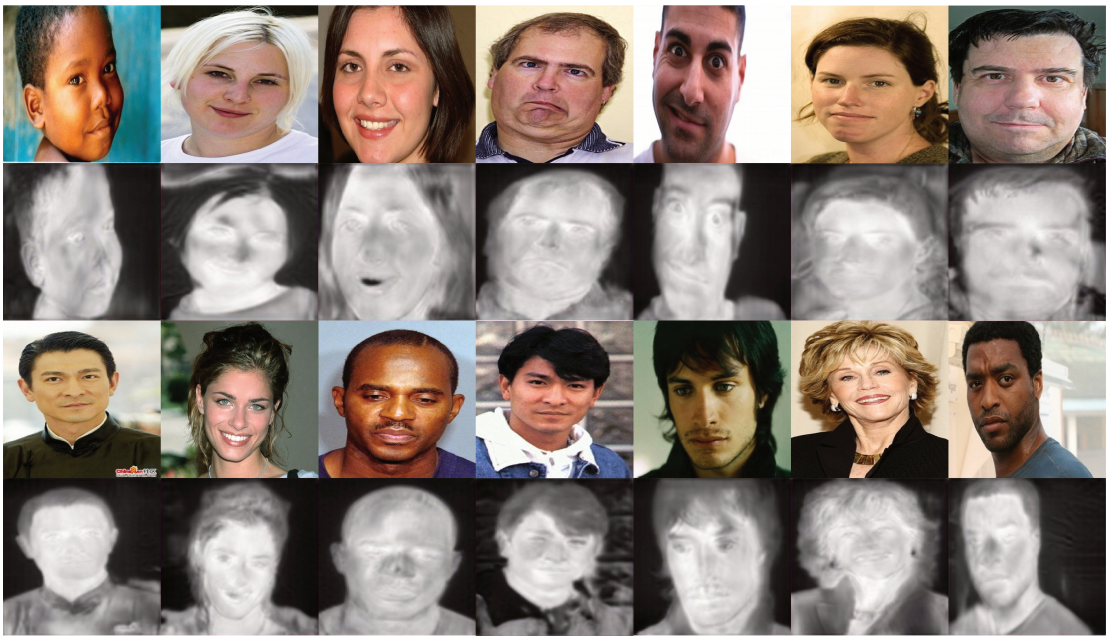


Figure 6.2: Samples of synthesized thermal images from HELEN and LFPW databases.

6.4 Facial landmark detection

In this section, we describe the two selected methods of FLD that will be trained on the synthesized thermal face databases.

6.4.1 Active appearance model

The first approach used in this work is based on Active Appearance Model [134], considered as the *baseline* approach for landmark detection. Active appearance models (AAM) were introduced by Cootes et al. [134] for facial image processing. AAM is a

statistical appearance method aiming to model the shape of the face and its appearance as probabilistic distributions that can be generalized nearly to any face. To train the FLD model, AAM requires a set of face images with annotation points defining the facial landmarks. In the training phase, Procrustes analysis [159] is applied to align the set of landmarks, and the statistical shape and appearance model variations are extracted using principal component analysis (PCA). Unseen faces can be represented by a linear combination of the mean shape and the appearance from the training data with weighted shape and appearance vector.

As to faithfully replicate the AAM approach used to train the FLD model provided by Aachen University [151], we have trained a dense histogram of gradients HOG feature-based AAM model fitted using the Inverse-Compositional algorithm [160].

6.4.2 Deep alignment network

The second selected approach is Deep Alignment Network (DAN) [135] as it is the state-of-the-art in facial landmark detection for visible images. DANs are based on multi-stage neural networks that perform an iterative process of refinement of landmark positions. Each stage of a DAN network is a feed-forward neural network that provides a prediction of the refined facial landmark location. Each stage of a DAN network takes 3 inputs: the original image aligned to an initial estimation of the landmark location, assumed to be the average face shape, the landmark heatmap, and the feature image provided by the previous stage. The first stage only takes the input image. The stages of DAN networks are trained consecutively. Each stage is trained until the validation error stabilises. We have used a two-stage DAN: between the two stages a similarity transform is applied to re-align the image to the average face shape. A learning rate of 1e-3 is used with Adam optimizer on mini batches of sizes 64.

6.5 Experimental setup and results

In this section, we present firstly our two baseline FLD models. Then, we detail our experimental setup. Finally, we introduce our evaluation protocol followed by a quantitative and qualitative evaluation on real thermal data of different quality.

6.5.1 Baseline models

We consider as baseline models the facial landmark detectors, described in Section 6.4, trained on high quality database provided by Kopaczka et al. [151] from University of Aachen. We will refer, in the remainder of this chapter, to the active appearance model

and deep alignment networks, both trained on Aachen database, as '*AAM-Aachen*' and '*DAN-Aachen*', respectively. The Aachen database includes high resolution thermal face images that are manually annotated [153]. Video sequences were acquired using a thermal camera with a NETD<30mK and spatial resolution of 1024×768 pixels. 695 frames were extracted and manually annotated with 68 point landmarks. To train the AAM model described in section 6.4.1, the face images were mirrored and 1272 images were selected for the training phase, as described in [151].

6.5.2 Experimental setup

The two selected approaches for FLD, described in section 6.4, are trained on the synthesized thermal face databases Helen and LFPW separately. We refer to AAM models trained on the synthesized thermal data from Helen and LFPW as ''*AAM-Helen*' and '*AAM-LFPW*' and to DAN models as '*DAN-Helen*' and '*DAN-LFPW*', respectively.

Following the protocol defined in the context of *300 Faces in-the-Wild Challenge: the first facial landmark localization Challenge [158]*, we have used 2000 face images from the Helen database and their corresponding facial landmark annotation files for training. Whereas for LFPW database, we have used 811 face images for training our models.

6.5.3 Evaluation protocol

The evaluation of FLD performance is assessed by comparing the estimated landmark coordinates to the ground truth. The normalized root mean square error (NRMSE), is computed, point-to-point, to assess the average localization error. NRMSE is considered as a standard metric to evaluate FLD performance [161] and it consists of the Euclidean distance between the predicted landmarks and the ground truth landmarks normalized by a predefined distance. Several normalization distances were defined for facial landmark detection evaluation [161, 162, 163, 164, 165]. To maintain consistency with the setup defined for the 300W competition [158], we performed the normalization with regards to inter-ocular distance (IOD) which is the distance between the two eye outer corners as defined in [158]. The normalization process is essential to obtain performance measurement independent of the face size or the image resolution.

The NRMSE, referred to as E, is obtained as follows:

$$E_k = \frac{\sqrt{((x,y)_k - (\bar{x},\bar{y})_k)^2}}{d_{norm}} \tag{6.1}$$

where $(x, y)_k$ denote the ground truth coordinates and $(\bar{x}, \bar{y})_k$ the estimated coordinates of the k^{th} landmark point. d_{norm} indicate the normalization distance.

The FLD performances can also be expressed in terms of detection rate. Facial landmark detection rate is the percentage of landmarks that are correctly detected within a given error radius. The accepted error radius is determined as a proportion of the IOD. The detection rate is calculated as follows:

$$D = \frac{\sum_{k=1}^{K} \sum_{i=1}^{N} [\delta : E_k^i \leq threshold]}{N \times K} \text{ , } where\ \delta = \begin{cases} 1 & \text{if } E_k^i \leq threshold \\ 0 & \text{otherwise} \end{cases} \tag{6.2}$$

where K denotes the total number of the facial landmarks, and N the number of test images. The threshold indicates the NRMSE value under which a landmark point is considered correctly localized. The IOD along with detection circles, representing the allowed error radius, are illustrated in Figure 6.3.

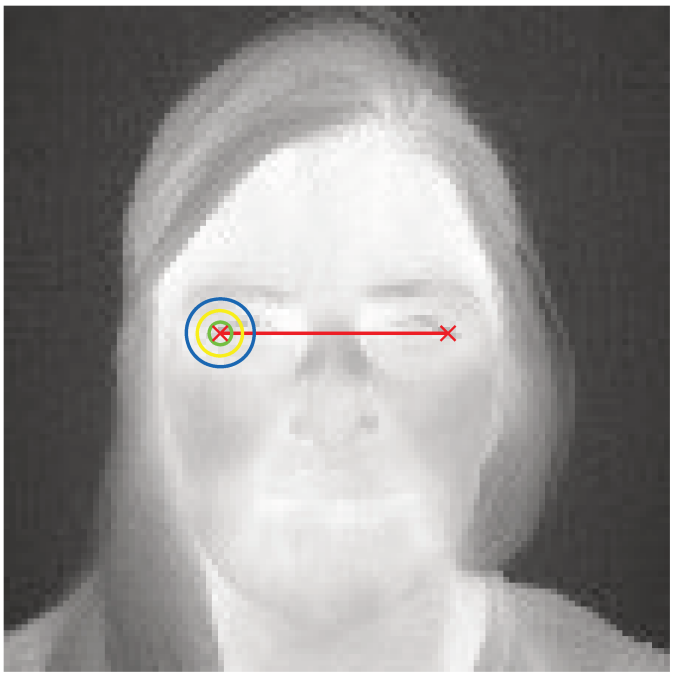


Figure 6.3: Inter-ocular distance (IOD) marked in red and circles denoting different detection error thresholds, green: 0.05, yellow: 0.1, blue: 0.15 times IOD.

6.5.4 Evaluation on low quality thermal face data

To evaluate the FLD model on low quality thermal data, the CSMAD database [51] is chosen, since it provides aligned images in visible and thermal spectrum acquired simultaneously. The CSMAD database provides thermal images of spatial resolution of 320×240 and NETD<70mK. This database is designed for face presentation attack, however, it is possible to select, for our evaluation, only the bona fide samples resulting in 423 images. The choice of this database is motivated by the fact that this database can simplify the annotation of the thermal images. The annotation process was performed automatically using DLIB [166] facial landmark detector on the visible set of the database and then corrected manually. Given that visible and thermal sets are aligned, the landmarks detected on the visible set are considered as the ground truth landmark points for the thermal set, as illustrated in Figure 6.4.

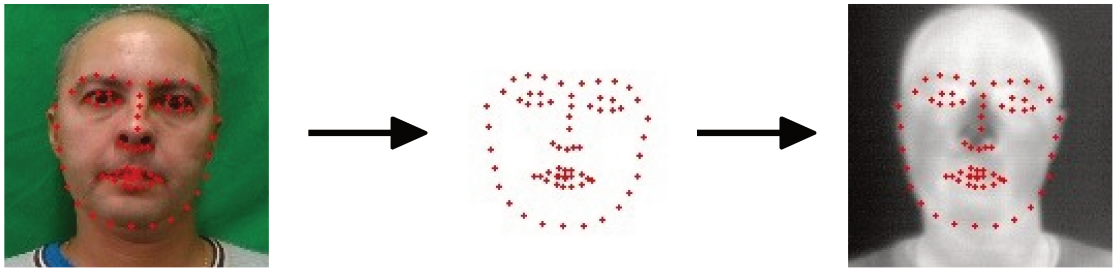
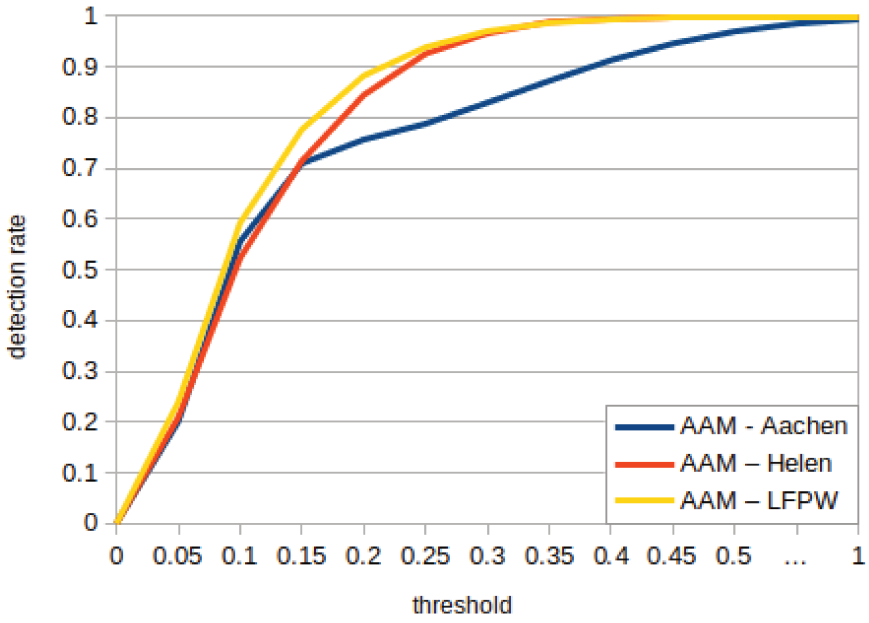


Figure 6.4: Ground truth facial landmark annotation of CSMAD data: facial landmarks are first detected on the visible images using DLIB [166] followed by manual verification and correction. The detected landmarks are simply used as ground truth for thermal images.

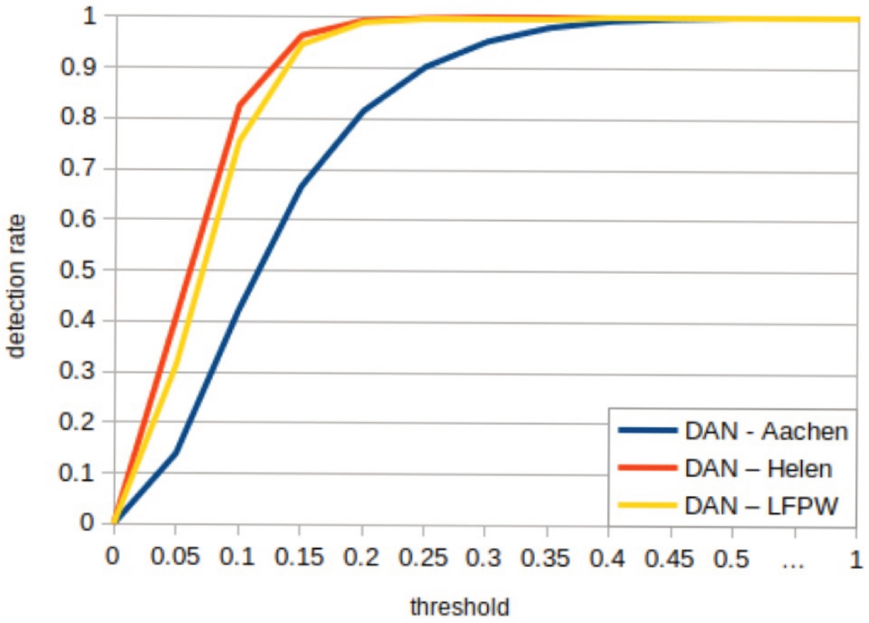
Given that this database also provides samples in visible spectrum, we trained the FLD approaches on the original visible face databases Helen and LFPW. FLD performance on the original visible database will be considered as a reference. The comparison of the performance obtained using a thermal based model with the visible based model will quantify the discrepancy of the two spectra in terms of FLD.

	AAM	**DAN**
Aachen (TH)	0.14349 (±0.105)	0.14595 (±0.052)
LFPW (SynTH)	0.11779 (±0.062)	0.08265 (±0.026)
Helen (SynTH)	0.13200 (±0.057)	0.07309 (±0.022)
LFPW (VIS)	0.04020 (±0.015)	0.04299 (±0.012)
Helen (VIS)	0.045683 (±0.031)	0.03146 (±0.011)

Table 6.2: Average NRMSE (± standard deviation) reported on CSMAD database.



(a)



(b)

Figure 6.5: Detection rate variation of facial landmark detection models evaluated on CSMAD database: (a) Active Appearance Model (b) Deep Alignment Network.

Results, in Table 6.2, show the average and the standard deviation of the localization error in terms of NRMSE obtained by evaluating the different FLD models on the CSMAD database. The first column of the table corresponds to the AAM approach trained on different databases: where '*TH*', '*SynTH*' and '*VIS*' refer to thermal data, synthesized thermal data and visible data, respectively. The second column reports the same results for a DAN-based approach. The localization errors reported by the FLD models trained and tested on thermal face data is relatively higher than the errors reported by the model trained and tested on the original visible images. This is mainly due to the conversion of the face images from highly informative domain, the visible spectrum, to a comparatively lower informative domain as the thermal spectrum, resulting in a loss of information relevant for accurate FLD. We also observe the detection models trained on synthesized thermal data exhibit considerably lower errors than the models trained on the Aachen database, which demonstrates the efficiency of our proposed solution. The reported results prove that a FLD model trained on synthesized thermal face data is more robust than a model trained on high quality thermal face data, and that is due to the large gap in data quality between the Aachen database [153] and the current existing thermal face databases.

The plots, presented in Figure 6.5, illustrate the detection rate that corresponds to a defined threshold value for FLD models trained on different databases. We swept the detection threshold from 0.0 to 1.0 with a step of 0.05. We observe that the two facial landmark detectors trained on the Aachen database, represented by the blue curve, led to significantly lower detection rates compared to the detectors trained on the synthesized thermal data. This can be justified by the fact that Aachen models have been trained on very high resolution, i.e. high contrast images captured with very high thermal sensitivity. These images are very different from the images provided by the publicly available thermal face databases, as it is the case for CSMAD database. In addition, the detection rates obtained using DANs are considerably higher than the detection rates obtained using AAM. This confirms the efficacy of deep learning solutions in the FLD task.

Additional qualitative results, presented in Figure 6.6, depict the performance of each model of FLD on thermal face images with some facial variations. We note that the facial landmark detectors trained on Aachen database [151], shown in column (c) and (f), fail to accurately localize most of the facial traits even under the least challenging variation. However, all the four models trained on the synthesized thermal data provide more accurate landmark localization. Furthermore, we observe that deep learning based detectors (columns (f), (g) and (h)) led to a more meticulous facial landmark localization compared to the statistical modelling based detector. Besides, deep learning models seem to be very robust against challenging facial variation such as occlusion by glasses (rows 2 and 4). These methods managed to predict the facial landmark coordinates that

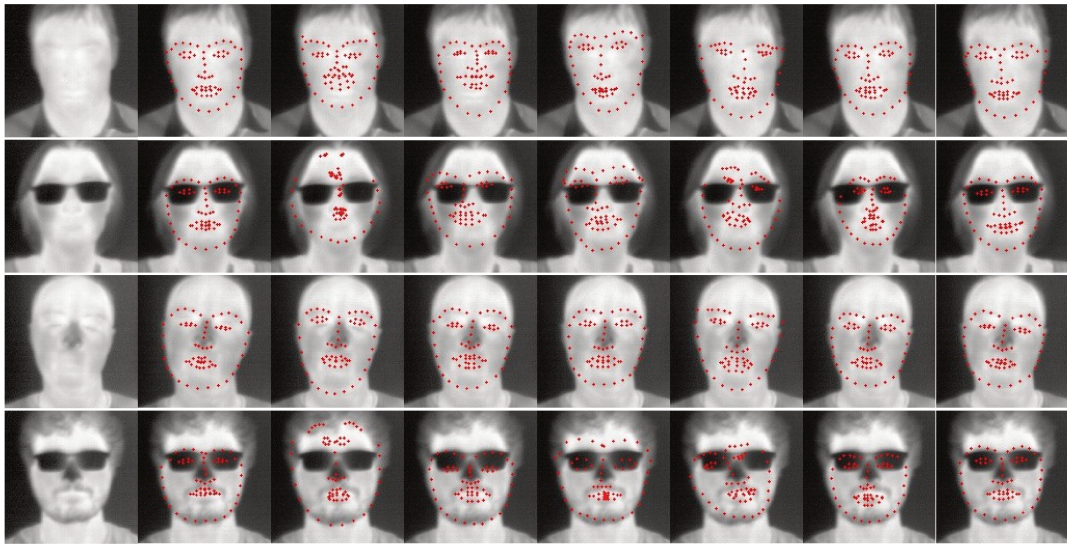


Figure 6.6: Qualitative results of the different facial landmark detection models on samples of CSMAD database.**(a)**: thermal reference, **(b)**: ground truth, **(c)**:AAM-Aachen, **(d)**: AAM-LFPW, **(e)**: AAM-Helen, **(f)**: DAN-Aachen, **(g)** DAN-LFPW, **(h)**: DAN-Helen.

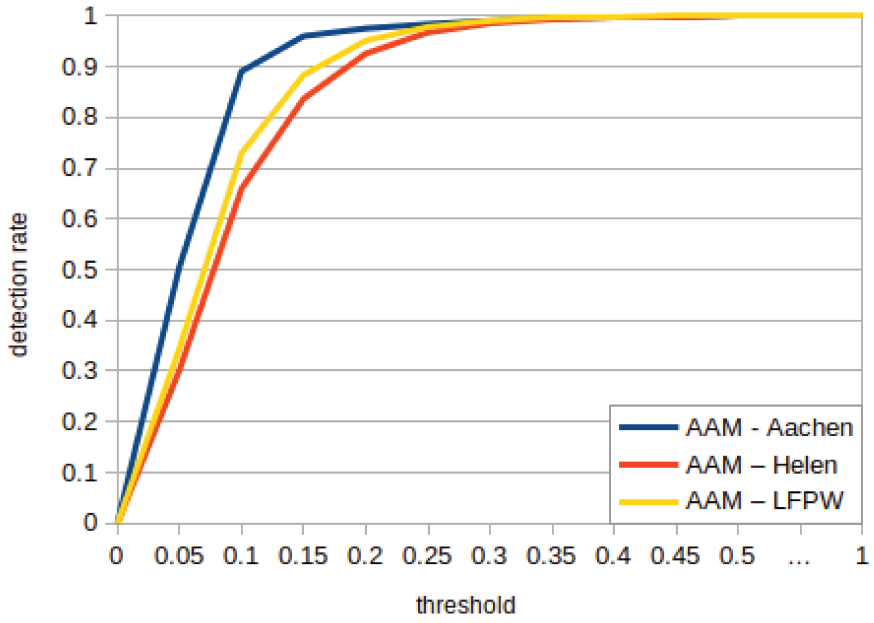
are closer to the ground truth, whereas the AAM based detectors tend to fail once it is tested on challenging face variations.

6.5.5 Evaluation on high quality thermal face data

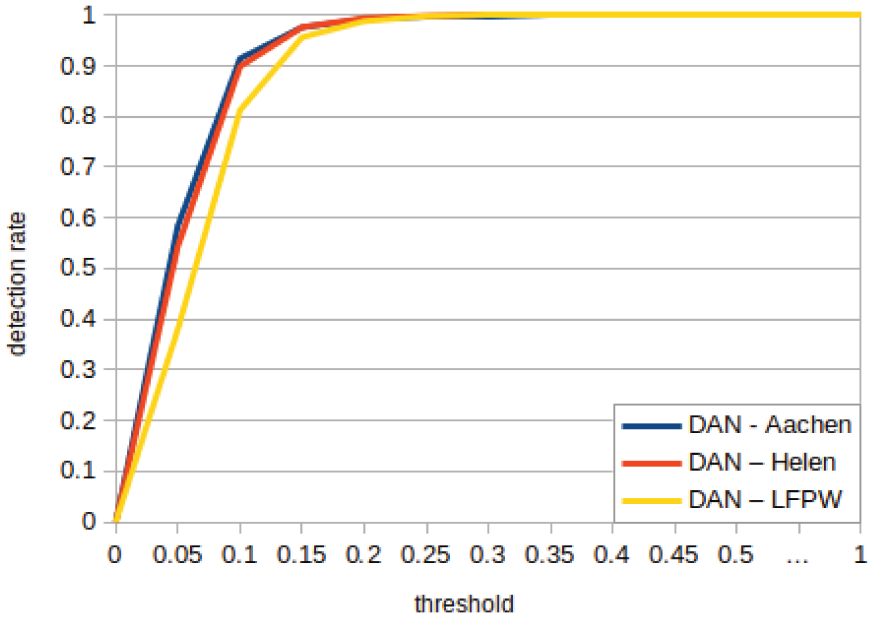
For fair comparison, the FLD models are also evaluated on high quality thermal data. The Aachen database [153] was extended to include thermal face images depicting facial expression variations providing 68 points landmark annotation as well. The expression variation subset of the Aachen database is used for our evaluation. Let us remind here that the data provided by the Aachen database is characterized by spatial resolution of 1024×768 pixels and NETD<30mK.

Table 6.3 presents the average and the standard deviation of the localization error of different FLD models when tested on the expression subset of the Aachen database. The detection models trained on Aachen database report lower, but with slight difference, localization errors than the detection models trained on synthesized thermal data. These results are somehow expected as the detection models trained on Aachen database are evaluated on data of the same thermal quality acquired with the same thermal sensor.

Detection rates of the different FLD models are illustrated in Figure 6.7. For the AAM



(a)



(b)

Figure 6.7: Detection rate variation of facial landmark detection models evaluated on the expression subset of the Aachen database: (a) active appearance model (AAM), (b) deep alignment network (DAN).

	AAM	DAN
Aachen (TH)	0.07267 (±0.031)	0.06061 (±0.020)
LFPW (SynTH)	0.09534 (±0.034)	0.07827 (±0.015)
Helen (SynTH)	0.10700 (±0.039)	0.06409 (±0.014)

Table 6.3: Average NRMSE (± standard deviation) reported on the expression subset of Aachen database.

approach, the detection rate reported by the model trained on Aachen data is significantly higher compared to the models trained on synthesized thermal data. However, for DAN, we notice that the curve corresponding to the model trained on the Aachen database overlaps with the curve obtained using the model trained on synthesized thermal data from Helen, attesting that the two models perform similarly.

Figure 6.8 presents some samples of the expression subset of Aachen database portraying the performance of each FLD model. Overall, FLD was less challenging when applied on high quality than on low quality thermal data, as revealed when we compare Figure 6.6 and Figure 6.8. For the AAM approach, facial landmark detectors trained on synthesized data perform slightly poorer than the detectors trained on the Aachen database. Nevertheless, when using DAN, the three different facial landmark detectors achieve similar performances as they all succeeded to meticulously locate the facial landmarks. For some face variations, we can observe that the model trained on the synthesized thermal Helen database (column (h)) detected adequately some challenging landmarks, as the bottom lip (row 1) and closed eyes (row 2), whereas the facial landmark detector trained on Aachen did not manage to correctly predict the localization of these landmarks (column (f)).

6.5.6 Qualitative evaluation on thermal samples of different quality

Given that there are no public thermal face databases, other than Aachen's [153], provided with full facial landmark annotation, further quantitative performance assessment cannot be performed on more data. Therefore, some qualitative results are illustrated in Figure 6.9 to demonstrate that the facial landmark detector trained on synthesized thermal data can operate accurately on thermal data of different quality. Results obtained using the DAN approach trained on Aachen database '*DAN-Aachen*' are shown in row 1 of Figure 6.9. We have presented, in row 3, results obtained using the DAN model trained on the synthesized thermal data from Helen database '*DAN-Helen*', as it is the best performing model.

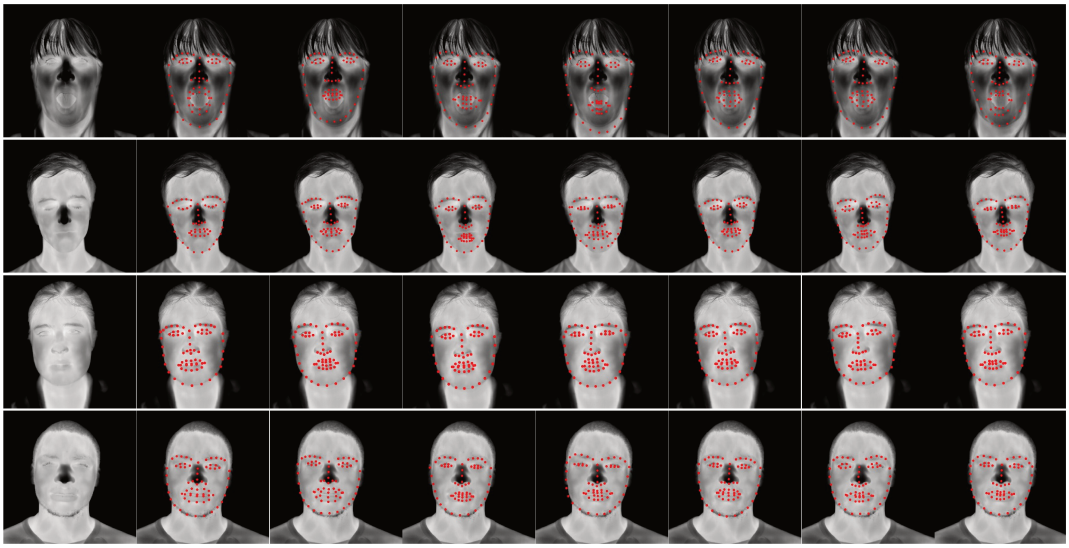
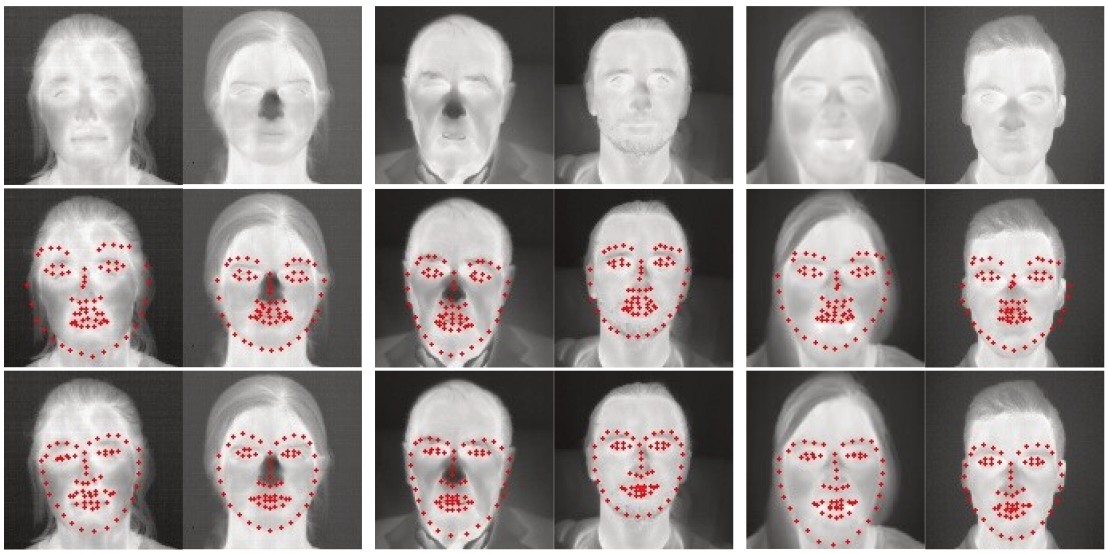


Figure 6.8: Qualitative results of the different facial landmark detection models on samples of the expression subset of Aachen database. **(a)**: thermal reference, **(b)**: ground truth , **(c)**:AAM-Aachen, **(d)**: AAM-LFPW, **(e)**: AAM-Helen, **(f)**: DAN-Aachen, **(g)** DAN-LFPW, **(h)**: DAN-Helen.

The presented samples are randomly selected from 3 different databases: (1) UND-X1 database [62, 63, 64] of spatial resolution of 312×239 pixels and NETD<100mK, (2) thermal face database provided by the Military University of Technology in Warsaw (UTW) [154] of spatial resolution of 640×480 and NETD<50mK, and (3) some samples from the high resolution version of VIS-TH database [105] acquired in our laboratory using a thermal sensor of spatial resolution of 620×512 and NETD<50mK. We can observe that for all the samples presented, the model trained on the synthesized thermal data '*DAN-Helen*' has succeeded to correctly localize the facial landmarks, outperforming the model trained on Aachen database '*DAN-Aachen*'.

Given all the results and observations presented above, one may conclude that our proposed concept has managed to obtain a facial landmark detector that can be suitable to a wide range of thermal image quality.



(a) (b) (c)

Figure 6.9: Qualitative results of facial landmark detection on samples of different thermal face databases, using DAN-Aachen in row 2 and DAN-Helen in row 3. **(a)**: UND-X1 database [62, 63, 64], **(b)**: thermal database of Military University of Technology in Warsaw (UTW) [154] **(c)**: samples from the High resolution version of VIS-TH database.

6.6 Summary

In this chapter, we addressed the lack of public thermal face databases provided with full annotation for face analysis applications. We introduced an unexplored concept consisting of converting data from one domain to another to tackle this shortage of annotated data. Particularly, we proposed to synthesize artificially a thermal face database with full landmark annotation by converting a existing face database in visible spectrum that have been designed for facial landmark detection task to thermal spectrum. Two different facial landmark approaches were trained on the synthesized thermal face data and tested on low quality and then on high quality thermal data, proving the robustness of the trained models. Our approach was evaluated and compared with two facial landmark detection baseline models provided by Kopazcka et al [151]. These baseline models were trained on high quality thermal data that led to a considerable decrease in performance when tested on thermal face databases that are publicly available. Conclusively, the facial landmark detection models trained on synthesized thermal data significantly outperformed the baseline models trained on Aachen database when evaluated on lower quality thermal data. Whereas, when tested on high quality thermal data, our proposed models perform similarly to the baseline models that is more adapted for thermal images of such quality.

The best performing model that we have trained on the synthesized thermal face data has achieved an average localization error of 0.07 and 94.59% of detection rate at threshold value of 0.15 when evaluated on low quality thermal data. This facial landmark detection model will be shortly made publicly available, as facial landmark detection is an essential step for many face analysis tasks and as of today there are no public facial landmark detection tools for thermal spectrum that are available. Interspectral data synthesis is also reproducible to tackle any lack of available data for tasks that requires extensive annotation.

Chapter 7

Indirect spoofing attack on thermal face biometric system

The robustness of thermal spectrum against spoofing attacks lies in the acquisition process of thermal properties by the thermal sensor. In this chapter, we propose a new type of attack on thermal face recognition systems, performed at post-sensor level. In 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, thermal face images are not easy to obtain, unlike visible face images that are abundantly available on the web. Therefore, we propose to generate synthetic thermal attacks by converting visible face images into thermal spectrum. To perform visible-to-thermal attack synthesis, we use the approach presented in Chapter 6 based on cascaded refinement networks (CRN) trained using contextual loss as described in Chapter 4. In a scenario where the imposter has prior knowledge about the spoofing countermeasure of the system, we introduce a new loss computed at local binary pattern (LBP) maps level to fool a LBP-based spoofing attack detection algorithm. The threat caused by the proposed attacks is then evaluated using two existing baselines of spoofing attack detection. The experimental results show that the new proposed attacks alter the performance of spoofing attack detection and lead to a higher error compared to the challenging presentation attack using silicone masks.

The remainder of this chapter is organized as follows. The context and motivation of this work are presented in Section 7.1. Section 7.2 presents the studies carried out for spoofing attacks on thermal spectrum. Section 7.3 recalls 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 7.4 details the process

to generate the proposed synthetic attacks and presents a quality assessment of the synthesized thermal images. Section 7.5 reports the experimental setup defined for the evaluation of two existing baselines of spoofing attack detection when confronted to the proposed attacks, followed by results and discussion. A summary is presented in Section 7.6.

7.1 Context and motivation

With the growing usage of face biometric systems, it is commonly acknowledged that this technology is exposed to multiple threats [167, 168, 169]. Eight different levels of attacks have been defined in [167, 170]. Considering exclusively the attacks that occur at biometric sample level, face biometric systems might be the most vulnerable among all other biometric systems, as faces are accessible on social networks or through capturing a photograph at a distance without the victim's consent. These attacks can be categorized, as illustrated in Figure 7.1, into: *direct* or *physical access* attacks, and *indirect* or *logical access* attacks.

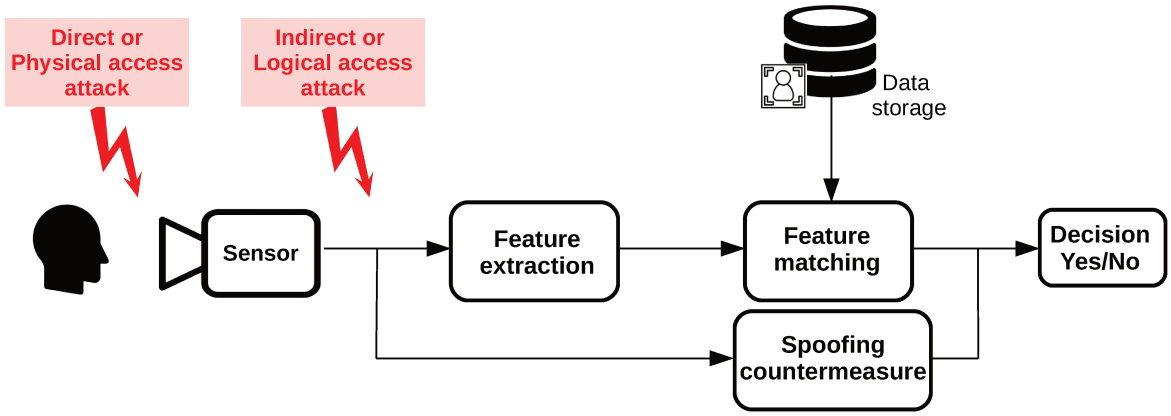


Figure 7.1: Attacks on biometric sample in a face biometric system.

Direct or physical access attacks occurs at pre-sensor level and are referred to as presentation attack. According to *ISO/IEC30107* standards [50], presentation attack is defined as *"the presentation of an artifact 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 unauthorized access, or to evade the biometric system by concealing the attacker's identity. The presented artifact can consist of a fake biometric sample of the claimed identity, e.g. photographs, masks, etc., in spoofing scenarios, or some alteration or falsification [56, 171] applied to the imposter's

own biometric sample in evasion scenarios. Recent research studies [51,52,53,55,56] have proved that using thermal imagery might be the most effective solution to presentation attack detection. The thermal signature of the human face provides evidence of the user's liveness. Artifacts presented by the imposter exhibit different thermal characteristics of those of a face, leading to a straightforward presentation attack detection solution.

Indirect or logical access attack, on the other hand, occurs at the post-sensor level. For this scenario, it is assumed that the impostor has access to the communication channel between the sensor and the feature extraction module, as shown in Figure 7.1. This kind of attack intercepts the face sample acquired by the sensor and substitutes it with a fake sample of the claimed identity. This attack can be as simple as inserting a photograph or replaying a video of the victim. Face samples are easy to obtain so as to spoof conventional visible spectrum based face biometric systems. However, this is not the case for thermal face biometric systems, as thermal images are not abundantly available.

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 that 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 it may encounter.

7.2 Literature overview: spoofing attacks and thermal spectrum

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 [172,173] merely 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 that present in thermal cameras, as it provides a proof of the user's liveness simply through acquisition [54]. 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 sensors based presentation attack detection. Thermal sensors remain highly robust against rigid 3D mask attacks, as the rigid mask presents a uniform pattern with much lower temperature than 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 [51, 52, 53] 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. Figure 7.2 depicts different types of spoofing attack. 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 [55, 56]. Face disguise can easily be detected since the used accessories present different thermal properties from those of a human face [56]. Thermal imagery can also identify plastic surgeries, as the resulted alteration of blood vessels appear as cold areas in the face [55].

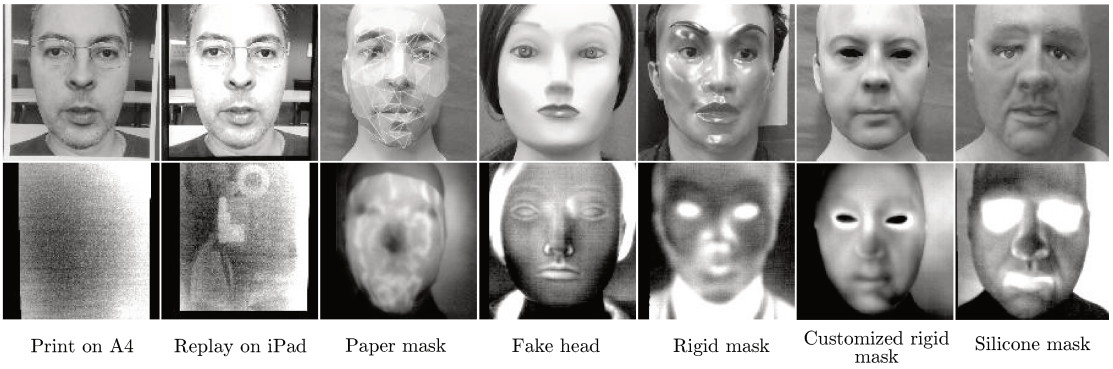


Figure 7.2: Presentation attacks in visible and thermal spectrum.

A preliminary study was carried out, by Bhattacharjee et al. [54], 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 be easily be detected in thermal spectrum by using the mean facial brightness of the face region. In [51], authors prove 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 brightness, as proposed in [54]. Agarwal et al. [53] 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 thermal spectrum is the most robust spectrum 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 machines (SVM). However, the results reported on thermal spectrum are questionable since the thermal data is clearly acquired using FLIR MSX* technology which adds visible light details to the thermal images. George et al. [52] present a new multi-channel database containing different 2D and 3D attacks. Multi-channel convolutional neural network (CNN) was proposed in this work for presentation attack detection. In addition, 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 [53]. In [56], a disguise database in visible and thermal spectrum was proposed. The authors proposed to combine patches from visible and thermal images for presentation attack detection.

7.3 Visible-to-thermal attack synthesis

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 reintroduce the used approach to convert visible images to thermal spectrum. A customization of the used approach is later presented to generate more challenging attacks to a given approach of thermal spectrum based presentation attack detection. Finally, implementation details of the proposed approaches are given.

7.3.1 Generalized approach for attack synthesis

Visible-to-thermal attack synthesis was carried out using the approach presented in Section 4.5 of Chapter 4. This approach is based on cascaded refinement networks (CRN) [21] trained using contextual loss [100]. In this case, the data synthesis is performed from visible to thermal spectrum as it is the case of Chapter 6 of this thesis. The synthesized attack to be generated is generalized to all spoofing attack detection algorithms. We reformulate the loss, defined in equation 4.5 in Chapter 4, to adapt it to visible-to-thermal image synthesis:

*FLIR MSX: https://www.flir.com/discover/professional-tools/what-is-msx/

$$\begin{aligned}\mathcal{L}_{CRN}(I_{VIS}, I_{TH}, G) =& \lambda_1(-\log(CX(\Phi^{l_s}(G(I_{VIS})), \Phi^{l_s}(I_{TH})))) + \\ & \lambda_2(-\log(CX(\Phi^{l_c}(G(I_{VIS})), \Phi^{l_c}(I_{VIS})))),\end{aligned} \tag{7.1}$$

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 VGG-19 embeddings extracted at content layers level and style layers level, respectively. CX denote the contextual similarity defined in Equation 4.4 in Chapter 4. λ_1 and λ_2 represent two emperically optimized weights associated to the style and content losses, respectively.

7.3.2 Customized approach for attack synthesis

Here, we explore the scenario in which an imposter has obtained prior information about the spoofing attack detection approach used in the targeted thermal face biometric system. Therefore, the generalized approach for attack synthesis will be customized according to this prior information.

The study, carried out by George et al. in [52], has proven that the spoofing attack detection algorithm based on LBP feature extraction is outperforming the solution provided by [53]. Therefore, we consider the LBP based spoofing attack detection as our target spoofing countermeasure on which the impostor has some prior information. Consequently, we customized our generalized visible-to-thermal attack synthesis 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 bona fide samples.

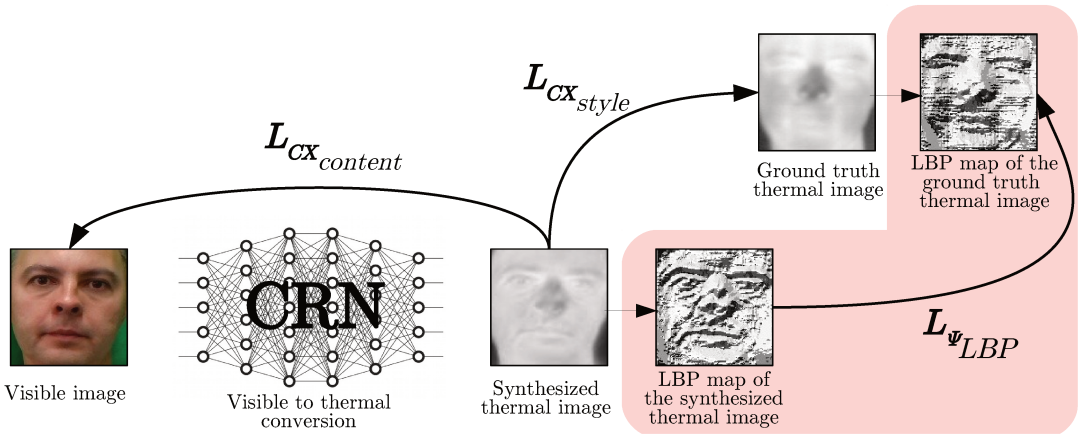


Figure 7.3: Diagram of the proposed approach to perform visible-to-thermal attack synthesis. The highlighted blocks of the diagram illustrate the introduced loss for the customized approach.

In Figure 7.3 the diagram of the different used approaches to perform visible-to-thermal attack synthesis is illustrated. Our LBP based customization of the CRN model is highlighted with an underlying light red area. The part of the diagram that is not highlighted, represents the generalized visible-to-thermal attack synthesis 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 (bona fide) image. In addition to the loss defined for the generalized visible-to-thermal attack synthesis model, we introduced a new loss that is computed at LBP map level. The LBP map is generated using a uniform pattern: 8 sample points in the neighborhood on the circle of radius 1. We propose to compute this loss function, denoted as Ψ in figure 7.3, in two different ways, as described in the following:

$\Psi=\chi^2$(LBP) The first option is to consider as loss function the LBP histograms comparison using χ^2 distance. The histogram of LBP labels is calculated over the whole LBP map, resulting in a feature vector of dimension 59. Training the visible-to-thermal attack synthesis network aims thus to minimize the χ^2 distance computed between the LBP histogram of the synthetic thermal image and the thermal ground truth image. The total loss of the customized attack synthesis model is formulated as follow:

$$\begin{aligned}\mathcal{L}_{Total}(I_{Vis}, I_{Th}, G) =& \alpha_1 \mathcal{L}_{CRN}(I_{Vis}, I_{Th}, G) + \\ & \alpha_2 \mathcal{L}_{\chi^2}(LBP_{hist}(G(I_{Vis})), LBP_{hist}(I_{Th}))\end{aligned} \tag{7.2}$$

Ψ=CX(LBP) The second option is to use a contextual loss computed on LBP maps, but solely at style level, as our objective is to generate thermal attacks of which the LBP maps is closer to the LBP map of thermal bona fide images. We extracted the VGG-19 embedding vectors from LBP maps of the synthetic thermal image and the thermal ground truth, at style layers. Consequently, the total loss of the customized attack synthesis model, in this case, is defined as follow:

$$\begin{aligned}\mathcal{L}_{Total}(I_{Vis}, I_{Th}, G) =& \alpha_1 \mathcal{L}_{CRN}(I_{Vis}, I_{Th}, G) + \\ & \alpha_2(-\log(CX(\Phi_{l_s}(LBP(G(I_{Vis}))), \Phi_{l_s}(LBP(I_{Th}))))\end{aligned} \tag{7.3}$$

In addition to the annotation defined in the equation 7.1, LBP and LBP_{hist} denote the LBP map and the histogram of the LBP map, respectively.

For the remainder of the paper, we refer to the visible-to-thermal attack synthesis models as *CRN*, *CRN+χ^2(LBP)*, and *CRN+CX(LBP)* to denote the generalized model, the customized model combined with LBP histogram comparison using χ^2 distance, and the customized model combined with the contextual loss at style level computed on LBP maps, respectively.

7.3.3 Implementation details

The different visible-to-thermal attack synthesis models are trained using the VIS-TH database presented in Chapter 3. One variation was discarded from the database, as it was acquired in total darkness. Visible and thermal images are re-sampled to 128×128 pixels.

The training of the three proposed models of visible-to-thermal attack synthesis was performed with a learning rate of 1e-4. The *CRN* model was run for 40 epochs, *CRN+χ^2(LBP)* model for 60 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.

7.4 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.

7.4.1 CSMAD dataset for indirect attack synthesis

Choosing the Custom Silicone Mask Attack Dataset (CSMAD) [51] 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 damage caused of the new attack, which we are proposing in this chapter, on spoofing attack detection will be quantified and compared to the damage brought by the silicone masks attack.

The CSMAD contains presentation attacks made of six custom-made silicone masks. Face images are collected from 14 subjects. Bona fide samples were collected from all subjects. Extra bona fide samples were acquired for few subjects, for which they wore eye glasses. 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 bona fide and attack acquisitions, consisting of videos of 5 to 10 seconds, in 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 visible and thermal spectrum. Figure 7.4 present some attack samples. We can observe, in column (a), when the mask is worn by the attacker it gets warm, leading to a thermal face sample that looks more like a real face in terms of temperature. 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.

7.4.2 Quality assessment of the synthetic attacks

Bona fide samples from the CSMAD dataset, that are acquired in visible spectrum, are simply fed to the visible-to-thermal attack synthesis models presented, in Section 7.3, 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 due to low illumination.

Figure 7.5 illustrates the synthetic attacks in column (c), (d) and (e). We note that the synthetic thermal images present realistic patterns of thermal signature. Some details, such as hair and eye brows, 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 different source of information. Comparing the synthetic thermal images generated using the three proposed visible-to-thermal attack

(a) (b)

Figure 7.4: Samples of presentation attack of CSMAD database in visible and thermal spectrum. **(a)** worn masks **(b)** standing masks.

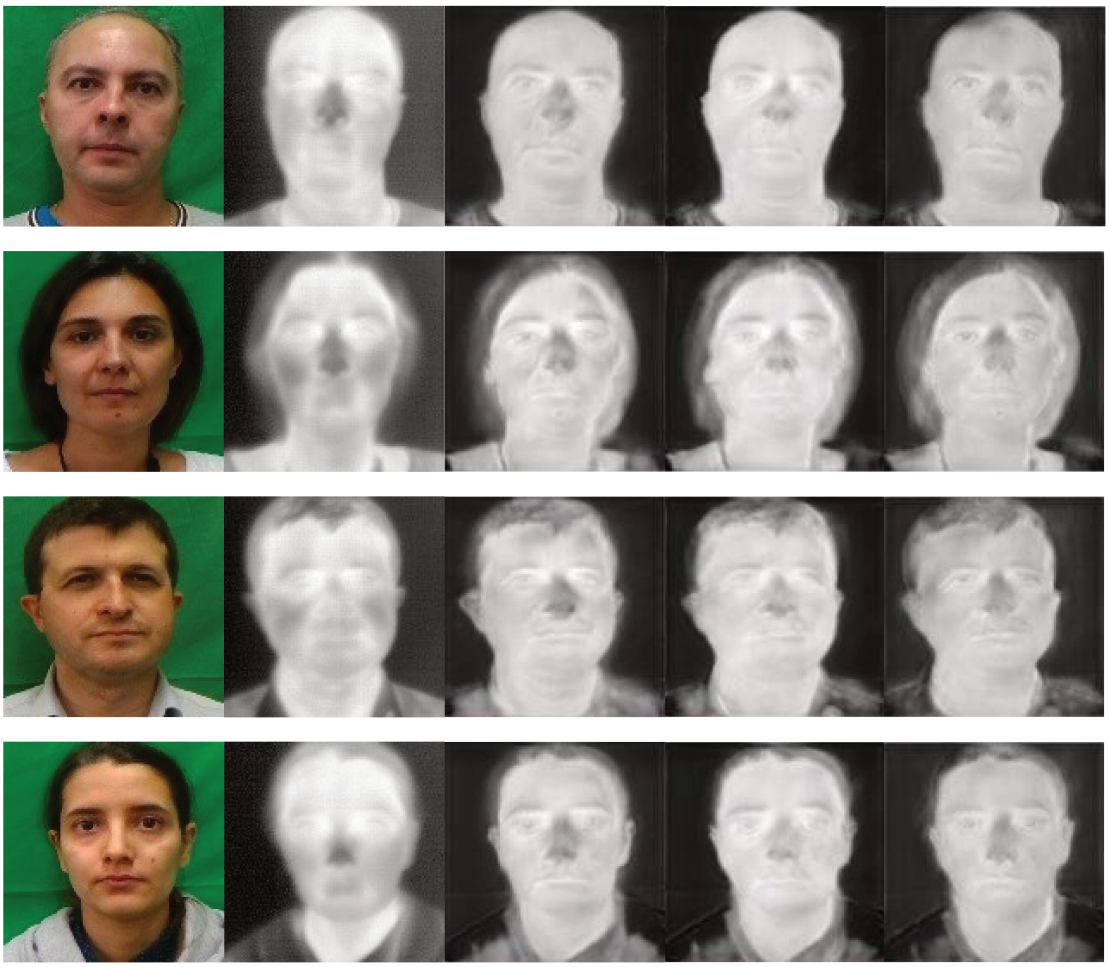


Figure 7.5: Samples of synthetic attacks. **(a)** visible bona fide **(b)** thermal bona fide **(c)** synthetic attacks using CRN **(d)** synthetic attacks using CRN+χ^2(LBP) **(e)** synthetic attacks using CRN+CX(LBP).

synthesis models, we note that the three sets of synthetic images are remarkably similar, even though we can note few minor differences that are almost not visually perceptible.

	PSNR (dB)	**SSIM**
CRN	15.576 (± 4.246)	0.610 (± 0.103)
CRN+χ^2(LBP)	15.223 (± 4.594)	0.613 (± 0.123)
CRN+CX(LBP)	15.616 (± 4.208)	0.618 (± 0.107)

Table 7.1: Quality assessment of the synthetic attacks in terms of PSNR and SSIM.

A quality assessment of the synthetic thermal attacks obtained by the different proposed approaches is performed in terms of peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM). PSNR and SSIM are computed between the synthetic thermal images and the thermal bona fide samples (ground truth). Table 7.1 reports the PSNR and SSIM results obtained for each visible-to-thermal attack synthesis model. We acknowledge that the obtained results do not reflect high fidelity of the synthetic thermal images to the ground truth. As pointed out for Figure 7.5, the synthetic attacks are generated from visible face images which provides a different information compared to thermal spectrum. The visible-to-thermal attack synthesis models aim to generate thermal-like images but it cannot predict accurately the thermal signature. The quality assessment provides similar results for the different attack synthesis models (∼15dB for PSNR and ∼0.6 for SSIM), with the *CRN+CX(LBP)* model delivering the highest values of PSNR and SSIM.

7.5 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 it causes. 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.

7.5.1 Spoofing attack detection baselines

The selected baselines of spoofing attack detection were introduced in studies of thermal spectrum robustness against spoofing attacks [51, 52, 53, 54].

Mean facial brightness (MFB) As defended in [51, 54], mean facial brightness is a simple but a very efficient solution to prove the user's liveness. This argument can be endorsed by the fact that face regions are rather bright in 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 face region even when worn by the attacker. Mean facial brightness can be used simply as spoofing attack detection score.

Local Binary Patterns and Logistic Regression (LBP+LR) Local binary patterns (LBP) are used to represent the texture variation between bona fide samples and attack samples. Subsequently, logistic regression (LR) is used to build a classifier to label samples as *bona fide* or *attack*. LBP features are normalized before training the LR model. We have applied normalization to zero mean and unit standard deviation using parameters extracted only from the bona fide feature set. Given a LR trained model, the output of this spoofing attack detection is the probability of a sample being a bona fide.

7.5.2 Experiments and results

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

Face regions are cropped by extracting the face coordinates on visible spectrum and projecting them on thermal face images. MFB is computed across the face region. Figure 7.6 illustrates the score distribution of MFB for bona fide samples and attack samples. The score distribution of bona fide samples is the same for all the Figures 7.6a, 7.6b, 7.6c and 7.6d, as we have considered the same bona fide set for the 4 sets of attacks. For the silicone mask attacks illustrated in Figure 7.6a, we observe that the two score distribution are clearly separated, resulting in a 2.3% of equal error rate (EER). However, the score distribution for the synthetic attack generated by the three different models of visible-to-thermal attack synthesis significantly overlaps with the score distribution of bona fide samples. The synthetic attack generated using CRN+χ^2(LBP) model gives the highest equal error rate of 67.7%. The EER reported on all of the three different synthetic attacks surpasses 50%. Accordingly, we can deduct that the proposed synthetic attack have led to a terrible failure of the spoofing attack detection solution based on MFB.

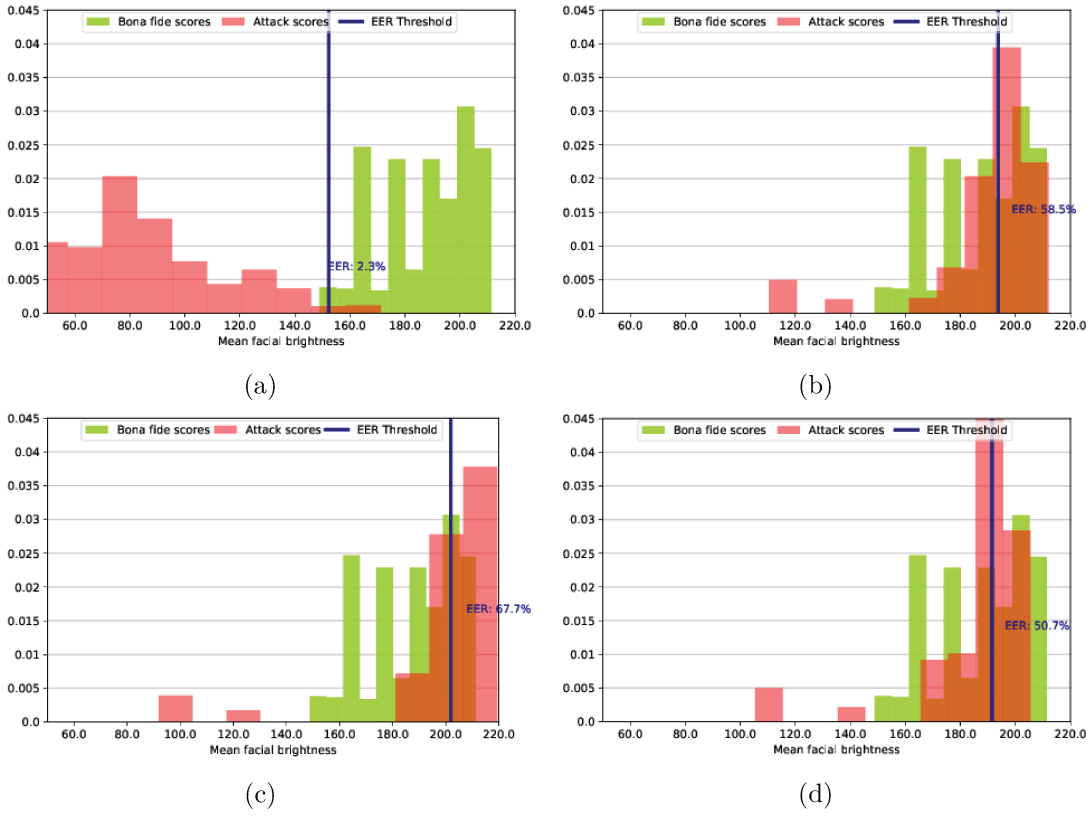


Figure 7.6: Score distribution of the MFB baseline for bona fide and attack samples. **(a)** silicone mask attack **(b)** synthetic attack CRN **(b)** synthetic attack CRN+χ^2(LBP), **(c)** synthetic attack CRN+CX(LBP)

For LBP+LR baseline, we split the CSMAD dataset into 14 partitions, each corresponding 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 way to ensure a disjoint set of subjects, so that the spoofing attack detection model does not learn subject-specific information. Figure 7.7 presents the detection error tradeoff (DET) curves corresponding to each of the studied attacks. For 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 LBP+LR baseline drastically decreases when dealing with the proposed synthetic attacks. In a scenario of extremely secure spoofing attack detection system where almost no impostor 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 three different models, we note that combining the CRN model with the loss computed at LBP map level led to more challenging attacks.

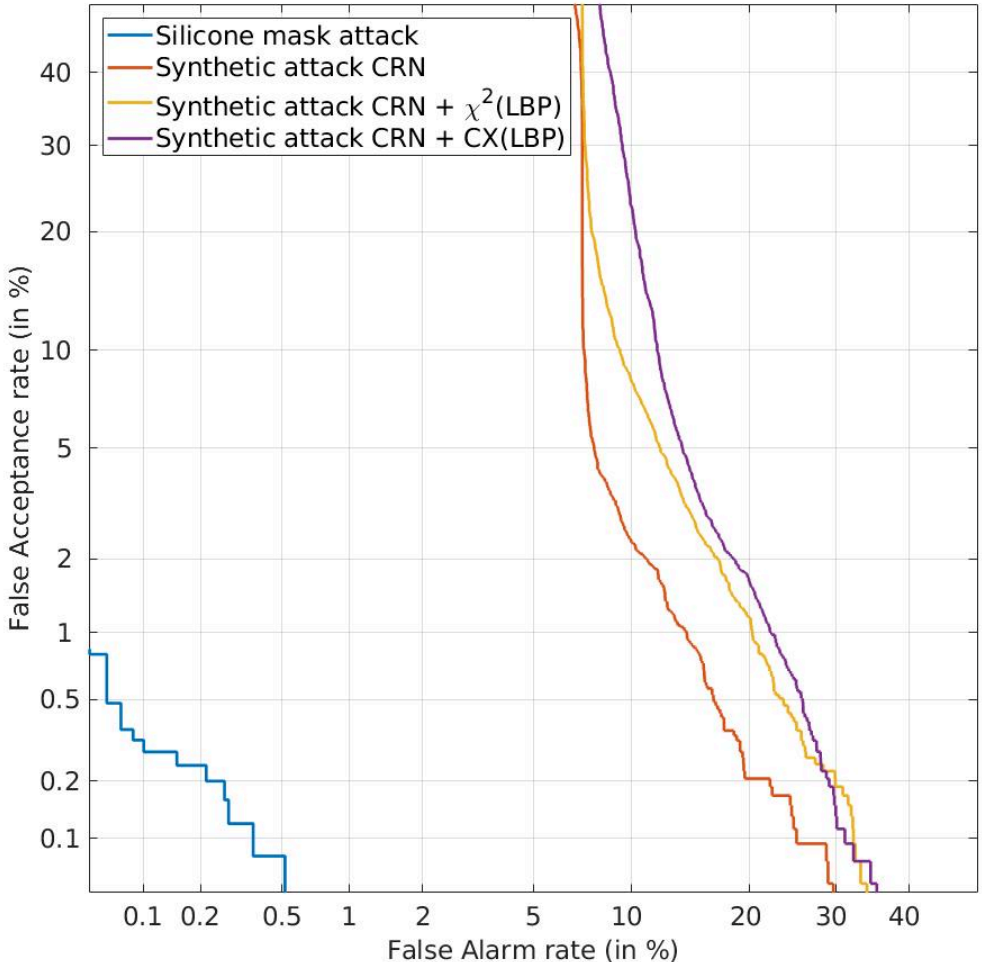


Figure 7.7: Detection error tradeoff (DET) curves of LBP+LR spoofing attack detection baseline for different attacks.

The EERs of the two reported spoofing attack baselines for the different attacks are gathered in Table 7.2. It is observable that the proposed synthetic attack represents a considerable higher threat, in comparison to silicone mask attack that is considered so far a challenging attack for thermal spectrum. The EER has increased from 2.3% to 67.7% and from 0.21% to 11.6% for MFB and LBP+LR spoofing attack detection, respectively.

	MFB	**LBP + LR**
Silicone mask attack	2.3	0.21
Synthetic attack CRN	58.5	7.43
Synthetic attack CRN + χ^2(LBP)	67.7	9.44
Synthetic attack CRN + CX(LBP)	50.7	11.6

Table 7.2: Equal error rate (%) of face spoofing attack detection evaluated on the proposed attacks.

When the impostor does not have any a priori knowledge about the spoofing countermeasure implemented in the system, the performance of the spoofing attack detection significantly drops when it faces the synthetic attack obtained by the generalized CRN model. Consequently the EER increased from 0.21% to 7.43%. Although when the impostor does indeed have a priori information about the spoofing countermeasure that is being employed, he can use this information in a way to customize his attack to have higher chances to breach the system. This scenario is executed for visible-to-thermal attack synthesis models, CRN + χ^2(LBP) and CRN + CX(LBP), where we have used the LBP map information to better attack the LBP+LR based spoofing attack detection system. In addition, it is important to highlight that when using a contextual loss at style level to compute the loss between the LBP maps of the synthetic thermal attack and the bona fide thermal sample, we have obtained a higher EER (11.6%) compared to using a LBP histogram comparison using χ^2 distance (9.44%).

7.6 Summary

Deploying thermal technology in face biometric systems requires an extensive study of its implications and the risk it may confront. In this chapter, 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 attack synthesis of visible face images that could be available on the social networks or acquired sneakily from a distance. A quality assessment of the synthetic attacks have been performed by comparing the synthesized thermal images to thermal bona fide samples. Subsequently, the threat of the proposed synthetic attack was measured through an evaluation of two existing spoofing attack detection solutions designed for thermal spectrum. This evaluation reported a significant drop of performance of the two used baselines when they face the proposed synthetic

attack compared to when they confront silicone mask attacks, the most challenging attack for thermal spectrum studied so far. A scenario representing an impostor that has a priori knowledge of the spoofing attack detection solution is also explored. For local binary pattern (LBP) based spoofing attack detection system, we have adjusted the visible-to-thermal attack synthesis model in a way that it aims to generate thermal images of which the LBP map is closer to the LBP map of thermal bona fide samples. The obtained synthetic attacks using the customised attack synthesis models have increased the error rate reported by the targeted spoofing attack detection approach.

We have proven through this work that, even though it is true that thermal spectrum is extremely robust against presentation attacks, this does not deny the fact that new attacks customized for thermal imagery might act as a serious threat. Spoofing attack detection approaches based on the detection of human vitals signs, such as respiratory rate or heart rate, might be an efficient, parallel, solution to counter-defend against the attacks proposed in this chapter.

Chapter 8

Conclusion

This chapter provides a summary of the contributions and findings from the work reported in this dissertation. This material is reported in Section 8.1. Different directions for future research are presented in Section 8.2.

8.1 Summary

Conventional visible face recognition systems have greatly evolved during the three last decades to achieve human-level performances. However, human performance does not always define an upper bound of what is achievable. Human vision system is limited by the potential of visible spectrum that detects reflected radiation in visible wavelengths. Thereby, visible face recognition systems are heavily affected by the illumination variation. Thermal imagery provides efficient solutions to the challenges encountered by visible face recognition systems. The foremost advantage of thermal imagery lies in its invariance to illumination changes. This is inherent in the nature of thermal imagery as it detects the radiation emitted by the face. Thermal face recognition has attracted a lot of attention these last years, however its progress is still far behind that of visible face recognition. This is mainly due to the shortage in thermal face databases and in public resources required for its exploration.

The research work reported in this thesis is centered on the development of novel methodologies that enable an efficient and prompt integration of thermal technology in face biometric systems. The set of developed methodologies, presented in this dissertation, was established based on interspectral synthesis that confers the exploitation of complementary information provided by face images in visible and thermal spectra. The proclivity for such direction is motivated by the explotion in usage of thermal technology

as the need and the investments for security applications grow steadily. The contributions presented throughout this thesis have promoted an integration of thermal technology without requiring:

- recollection of face enrollment databases in thermal spectrum as the legacy enrollment databases are restrained to visible spectrum.
- adapted and re-optimized algorithms specifically designed for thermal spectrum.
- extensive manual annotation and labeling of thermal data that is costly and time-consuming.

The shortage of public face databases that provide face images in visible and in thermal spectrum has motivated the first contribution of our work. A new face database, introduced in **Chapter 2**, includes face images acquired simultaneously in visible and in thermal spectrum using a dual sensor. The proposed database has been acquired with several facial variations in attempt to reproduce real-life challenging scenarios. Because of its variation, this database can be used to conduct a wide range of studies related to facial image processing including occlusion removal, expression and/or pose invariant face recognition and soft biometrics. A benchmark evaluation of the database has been conducted to study the impact of facial variations on visible and on thermal face recognition performance validating the advantages and the limitations of each. The database has been available upon request for the research community. The remainder of the contributions reported in this dissertation are built upon the representations provided by the proposed database.

The contribution, introduced in **Chapter 4**, relates to our fist application of interspectral synthesis and that is to perform cross-spectrum face recognition. Thermal-to-visible image synthesis is based on cascaded neural network (CRN) [21]. The training of CRN was performed using contextual loss [100] that enabled a scale and rotation invariant transformation. The proposed approach was, qualitatively and quantitatively, evaluated and compared to the state-of-the-art approach in image translation, Pix2Pix [96] and to a thermal-to-visible synthesis approach based on generative adversarial networks, TV-GAN [84], designed for cross-spectrum face recognition. The experimental results revealed the efficiency of our approach in bridging the gap between thermal and visible spectrum compared to the TV-GAN baselines by reporting an average of 56% of relative improvement in terms of face recognition accuracy. The presented contribution enables the straightforward integration of thermal technology in deployed face recognition systems without the need of recollection of face enrollment data in thermal spectrum, neither the re-configuration of inner processing modules designed for visible spectrum.

The work, presented in Chapter 4, was then extended to develop an illumination-invariant face recognition system using visible and thermal-to-visible face images. **Chapter 5** introduced a new scheme of score level fusion that leverages the more informative spectrum in given illumination conditions, yielding to a continuous day and night face recognition. While the reported results in Chapter 4 proved the efficacy of the thermal-to-visible image synthesis, the quality of the synthesized visible images are still few steps behind standard visible images. Based on the intuition that the quality of a sample can be an indicator of its relevance in providing an accurate recognition, the matching scores of visible images and thermal-to-visible images against visible gallery are associated with a quality matching score that compares the quality of the probe sample to the gallery sample. The proposed fusion scheme was employed in two face recognition systems, the first based on handcrafted features, i.e. local binary patterns [115], and the second based on deep neural embeddings extracted using LightCNN model [113]. The experimental results validate our approach as slight improvements in face recognition accuracy were reported .

The contribution of the work presented in **Chapter 6** consists in introducing a novel concept, that to our knowledge has not been previously explored, aiming to tackle the lack of annotated data in domains, other than visible spectrum, that are less studied in the field of image processing. The proposed solution consists of transferring the data from one domain, generally visible spectrum, to a target domain and using the converted data along with the original annotation to train a model designed to perform a determined task. Particularly in this dissertation, we have considered thermal spectrum as our target domain and facial landmark detection as the task to be performed. The data synthesis method has been adapted to perform visible-to-thermal data transformation. Two facial landmark detection methods, the first based on active appearance models [134] and the second based on deep learning technique [135], were trained on the synthesized thermal databases using the corresponding annotation. The evaluation results have reported a 44% of relative improvement in terms of accuracy detection over the baseline system.

Chapter 7 presents a new attack on biometric samples at the post-sensor level for thermal face biometric systems. These systems were proved to be very robust against spoofing attacks, however this robustness lies in the process of acquisition characterizing thermal sensors by detecting the thermal signature of the face. Therefore, the indirect access attacks, that occur at the post-sensor level, are an irrefutable threat that jeopardize the security granted by thermal face biometric systems. It is presumed 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, to the best of our knowledge, has not yet been explored in literature. Since thermal face images are nearly impossible to obtain, the proposed new attack consists of generating synthetic

thermal face images by transforming images acquired in visible spectrum to thermal spectrum. The scenario, where the impostor has a priori knowledge about the spoofing countermeasure used in the system and uses this information to adapt his synthetic attack to better spoof the system, is also considered. The threat of the proposed synthetic attacks is quantified using existing countermeasure approaches designed for thermal spectrum. The experimental results of spoofing attack detection show a relative increase in terms of equal error rate from 0.21% for silicone mask attack to 11.6% for the proposed synthetic attack demonstrating the risk it generates.

8.2 Directions for future research

Directions for future research relate to both the extension of the presented work for other facial image processing tasks as well as the generalization of the proposed methods for further computer vision applications. Further works include:

- **High resolution face paired database in visible and thermal spectrum** As stated in Chapter 3, a high resolution version of the database introduced in this thesis is being collected. This version of the database is being acquired with FLIR DUO PRO R sensor, that provides visible images of spatial resolution of 4000×3000 and thermal images of spatial resolution of 640×512 and thermal sensitivity lower than 50mK. In addition to the variations considered for the first version of the database, a variety of metadata is also being collected that includes weight, height and wrist size that will lay the ground to explore the possibility of body measurements estimation from face images. The database will also provide a 1 minute long face videos along with the measurement of heart rate. This will enable monitoring cardiorespiratory signals using thermal faces. The collection of this high resolution database is essential for the research community to keep up with the rapid advancements of thermal imaging technology.

- **Sppofing countermeasure for indirect spoofing attack on thermal biometric systems** Following the last contribution of this thesis presented in Chapter 7, a spoofing detection solution can be proposed in thermal spectrum based on the extraction of subcutaneous information that the thermal face images provide. One possible direction is the extraction of cardiac signals to prove the user's liveness. The new database collection will provide the data required for the development of such countermeasure technique. Another solution can be based on the usage of subcutaneous information provided by the thermal images. Thermal face recognition relying on the extraction of subcutaneous features such as vascular network matching [174] or blood perfusion data [175] can be directly employed.

- **Improvement of interspectral face synthesis** While the interspectral face synthesis used for cross-spectrum face recognition yielded to a significant improvement compared to the baseline systems, the synthesized visible face images still present few artefacts when the face is presented under challenging face variations such as head pose and occlusions. Other artefacts are related to incorrect estimation of some facial attributes such as gender and skin color. Improvements will be explored with the aim of addressing the aforementioned artefacts to provide higher cross-spectrum face recognition accuracy and enhanced quality face images.

- **Application of interspectral synthesis for crowd density estimation** The research work reported in this dissertation has been already proved to be a low-hanging fruit. New projects have started to be proposed basing their research scope on interspectral image synthesis for applications other than that of facial image processing. An ongoing project entitled "*OKLOS: Continuous anomaly detection in moving crowds*"* is drawing its focus on applying thermal-to-visible image synthesis for video surveillance tasks. This project has been selected by the French research agency (ANR) in the context of ANR Flash Call for Project: "*Security of the 2024 Olympic & Paralympic Games*". Thermal-to-visible image synthesis will lay the foundation for continuous day and night monitoring and surveillance, by means of the wide range of available resources in the visible spectrum. These resources include crowd motion analysis, density detection, and group behavior analysis.

*OKLOS website: http://oklos.eurecom.fr/

Bibliography

[1] S. Mitra and M. Gofman, *Biometrics in a Data Driven World: Trends, Technologies, and Challenges*. CRC Press, 2016. [Cited on pages 1].

[2] H. Han, S. Shan, X. Chen, and W. Gao, "A comparative study on illumination preprocessing in face recognition," *Pattern Recognition*, vol. 46, no. 6, pp. 1691–1699, 2013. [Cited on pages 2].

[3] X. Zou, J. Kittler, and K. Messer, "Illumination invariant face recognition: A survey." IEEE International Conference on Biometrics: Theory, Applications, and Systems, 2007. [Cited on pages 2].

[4] D. A. Socolinsky and A. Selinger, "A comparative analysis of face recognition performance with visible and thermal infrared imagery." Proc. International Conference on Pattern Recognition, 2002. [Cited on pages 2, 16, 30].

[5] X. Zhang and Y. Gao, "Face recognition across pose: A review," *Pattern Recognition*, vol. 42, no. 11, pp. 2876–2896, 2009. [Cited on pages 2].

[6] T. Jiang, T. Wang, B. Ding, and H. Wu, "Degan: De-expression generative adversarial network for expression-invariant face recognition by robot vision," in *2019 WRC Symposium on Advanced Robotics and Automation (WRC SARA)*. IEEE, 2019, pp. 209–214. [Cited on pages 2].

[7] S. Wang and Y. Fu, "Face behind makeup," in *Thirtieth AAAI Conference on Artificial Intelligence*, 2016. [Cited on pages 2].

[8] A. Hadid, "Face biometrics under spoofing attacks: Vulnerabilities, countermeasures, open issues, and research directions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2014, pp. 113–118. [Cited on pages 2].

[9] N. Kose and J.-L. Dugelay, "On the vulnerability of face recognition systems to spoofing mask attacks," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2013, pp. 2357–2361. [Cited on pages 2].

[10] L. Li, P. L. Correia, and A. Hadid, "Face recognition under spoofing attacks: countermeasures and research directions," *IET Biometrics*, vol. 7, no. 1, pp. 3–14, 2017. [Cited on pages 2].

[11] R. Ramachandra and C. Busch, "Presentation attack detection methods for face recognition systems: A comprehensive survey," *ACM Computing Surveys (CSUR)*, vol. 50, no. 1, pp. 1–37, 2017. [Cited on pages 2].

[12] J. Galbally, S. Marcel, and J. Fierrez, "Biometric antispoofing methods: A survey in face recognition," *IEEE Access*, vol. 2, pp. 1530–1552, 2014. [Cited on pages 2].

[13] R. Tolosana, R. Vera-Rodriguez, J. Fierrez, A. Morales, and J. Ortega-Garcia, "Deepfakes and beyond: A survey of face manipulation and fake detection," *arXiv preprint arXiv:2001.00179*, 2020. [Cited on pages 2].

[14] L. Souza, L. Oliveira, M. Pamplona, and J. Papa, "How far did we get in face spoofing detection?" *Engineering Applications of Artificial Intelligence*, vol. 72, pp. 368–381, 2018. [Cited on pages 2].

[15] M. Rai, T. Maity, and R. K. Yadav, "Thermal imaging system and its real time applications: a survey," *Journal of Engineering Technology*, pp. 290–303, 2017. [Cited on pages 2, 38].

[16] R. S. Ghiass, O. Arandjelović, A. Bendada, and X. Maldague, "Infrared face recognition: A comprehensive review of methodologies and databases," *Pattern Recognition*, vol. 47, no. 9, pp. 2807–2824, 2014. [Cited on pages vii, 2, 11, 13, 21].

[17] Texas instruments - 1966 first flir units produced. [Online]. Available: ti.com [Cited on pages 2].

[18] C. Hazbun, "The history of thermal imaging cameras," *ECAMSECURE*, February 2018. [Cited on pages 2].

[19] (2018, February) The next generation thermal by flir smartphone: Cat s61. [Online]. Available: https://www.flir.com/oem/thermal-by-flir/ [Cited on pages 3].

[20] (2016, July) Cat s60 review: A rugged phone that can see in the dark. [Online]. Available: https://www.theverge.com/circuitbreaker/2016/7/11/12147948/cat-s60-review-thermal-camera-waterproof-phone [Cited on pages 3].

[21] Q. Chen and V. Koltun, "Photographic image synthesis with cascaded refinement networks," in *IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy, October 22-29*. IEEE Computer Society, 2017, pp. 1520–1529. [Online]. Available: https://doi.org/10.1109/ICCV.2017.168 [Cited on pages 4, 40, 41, 99, 114].

[22] R. Munir and R. A. Khan, "An extensive review on spectral imaging in biometric systems: Challenges & advancements," *Journal of Visual Communication and Image Representation*, vol. 65, p. 102660, 2019. No citations.

[23] A. A. Richards, "Alien vision: exploring the electromagnetic spectrum with imaging technology." SPIE, 2011. [Cited on pages 11].

[24] M. A. Akhloufi and A. Bendada, "Fusion of active and passive infrared images for face recognition," in *Thermosense: Thermal Infrared Applications XXXV*, vol. 8705. International Society for Optics and Photonics, 2013, p. 87050B. [Cited on pages vii, 12].

[25] International commision on illumination. Http://www.cie.co.at/. [Cited on pages xi, 13].

[26] X. Maldague *et al.*, "Theory and practice of infrared technology for nondestructive testing," 2001. [Cited on pages 13].

[27] S. Hu, J. Choi, A. L. Chan, and W. R. Schwartz, "Thermal-to-visible face recognition using partial least squares," *J. Opt. Soc. Am. A*, vol. 32, no. 3, pp. 431–442, Mar 2015. [Online]. Available: http://josaa.osa.org/abstract.cfm?URI=josaa-32-3-431 [Cited on pages vii, 14, 38].

[28] G. Friedrich and Y. Yeshurun, "Seeing people in the dark: Face recognition in infrared images," in *International Workshop on Biologically Motivated Computer Vision*. Springer, 2002, pp. 348–359. [Cited on pages 14, 15, 17].

[29] A. Cheng-bin and W. Ying, "Analysis of netd test for thermal imaging system [j]," *Infrared and Laser Engineering*, vol. 3, 2010. [Cited on pages 14].

[30] D. A. Socolinsky and A. Selinger, "Thermal face recognition in an operational scenario," in *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*, vol. 2. IEEE, 2004, pp. II–II. [Cited on pages 15, 61].

[31] M. K. Bhowmik, D. Bhattacharjee, M. Nasipuri, D. K. Basu, and M. Kundu, "Optimum fusion of visual and thermal face images for recognition," in *2010 Sixth International Conference on Information Assurance and Security*. IEEE, 2010, pp. 311–316. [Cited on pages 15].

[32] O. Arandjelovic and R. Cipolla, "A new look at filtering techniques for illumination invariance in automatic face recognition," in *7th International Conference on Automatic Face and Gesture Recognition (FGR06).* IEEE, 2006, pp. 449–454. [Cited on pages 15].

[33] O. Arandjelovic, R. Hammoud, and R. Cipolla, "On person authentication by fusing visual and thermal face biometrics," in *2006 IEEE International Conference on Video and Signal Based Surveillance.* IEEE, 2006, pp. 50–50. [Cited on pages 15, 61].

[34] O. Arandjelović, R. Hammoud, and R. Cipolla, "Thermal and reflectance based personal identification methodology under variable illumination," *Pattern Recognition*, vol. 43, no. 5, pp. 1801–1813, 2010. [Cited on pages 15, 61].

[35] S. Moon, S. G. Kong, J.-H. Yoo, and K. Chung, "Face recognition with multiscale data fusion of visible and thermal images," in *2006 IEEE International Conference on Computational Intelligence for Homeland Security and Personal Safety.* IEEE, 2006, pp. 24–27. [Cited on pages 15, 16].

[36] O.-K. Kwon and S. G. Kong, "Multiscale fusion of visual and thermal images for robust face recognition," in *CIHSPS 2005. Proceedings of the 2005 IEEE International Conference on Computational Intelligence for Homeland Security and Personal Safety, 2005.* IEEE, 2005, pp. 112–116. [Cited on pages 15, 16].

[37] E. Zahran, A. Abbas, M. Dessouky, M. Ashour, and K. Sharshar, "High performance face recognition using pca and zm on fused lwir and visible images on the wavelet domain," in *2009 International Conference on Computer Engineering & Systems.* IEEE, 2009, pp. 449–454. [Cited on pages 15, 16].

[38] S. G. Kong, J. Heo, F. Boughorbel, Y. Zheng, B. R. Abidi, A. Koschan, M. Yi, and M. A. Abidi, "Multiscale fusion of visible and thermal ir images for illumination-invariant face recognition," *International Journal of Computer Vision*, vol. 71, no. 2, pp. 215–233, 2007. [Cited on pages 16, 18].

[39] H. Hariharan, A. Koschan, B. Abidi, A. Gribok, and M. Abidi, "Fusion of visible and infrared images using empirical mode decomposition to improve face recognition," in *2006 International Conference on Image Processing.* IEEE, 2006, pp. 2049–2052. [Cited on pages 16, 61].

[40] B. Abidi, S. Huq, and M. Abidi, "Fusion of visual, thermal, and range as a solution to illumination and pose restrictions in face recognition," in *38th Annual 2004 International Carnahan Conference on Security Technology, 2004.* IEEE, 2004, pp. 325–330. [Cited on pages 17].

[41] F. M. Pop, M. Gordan, C. Florea, and A. Vlaicu, "Fusion based approach for thermal and visible face recognition under pose and expresivity variation," in *9th RoEduNet IEEE international conference.* IEEE, 2010, pp. 61–66. [Cited on pages 17].

[42] N. Zaeri, "Pose invariant thermal face recognition using ami moments," in *2016 UKSim-AMSS 18th International Conference on Computer Modelling and Simulation (UKSim).* IEEE, 2016, pp. 60–64. [Cited on pages 17].

[43] A. Kwasniewska, J. Ruminski, M. Szankin, and K. Czuszynski, "Pose-invariant face detection by replacing deep neurons with capsules for thermal imagery in telemedicine," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC).* IEEE, 2018, pp. 778–781. [Cited on pages 17].

[44] S. Joardar, D. Sen, D. Sen, A. Sanyal, and A. Chatterjee, "Pose invariant thermal face recognition using patch-wise self-similarity features," in *2017 Third International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN).* IEEE, 2017, pp. 203–207. [Cited on pages 17].

[45] P. Buddharaju, I. T. Pavlidis, and P. Tsiamyrtzis, "Pose-invariant physiological face recognition in the thermal infrared spectrum," in *2006 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'06).* IEEE, 2006, pp. 53–53. [Cited on pages 17, 38].

[46] A. Gyaourova, G. Bebis, and I. Pavlidis, "Fusion of infrared and visible images for face recognition," in *European Conference on Computer Vision.* Springer, 2004, pp. 456–468. [Cited on pages 17, 60].

[47] S. Singh, A. Gyaourova, G. Bebis, and I. Pavlidis, "Infrared and visible image fusion for face recognition," in *Biometric Technology for Human Identification*, vol. 5404. International Society for Optics and Photonics, 2004, pp. 585–596. [Cited on pages 17, 60].

[48] J. Heo, S. G. Kong, B. R. Abidi, and M. A. Abidi, "Fusion of visual and thermal signatures with eyeglass removal for robust face recognition," in *2004 Conference on Computer Vision and Pattern Recognition Workshop.* IEEE, 2004, pp. 122–122. [Cited on pages 18, 61].

[49] W. K. Wong and H. Zhao, "Eyeglasses removal of thermal image based on visible information," *Information Fusion*, vol. 14, no. 2, pp. 163–176, 2013. [Cited on pages 18].

[50] I. 30107, "Information technology biometric presentation attack detection part 1: Framework," *International Organization for Standardization*, Jan 2016. [Cited on pages 18, 96].

[51] S. Bhattacharjee, A. Mohammadi, and S. Marcel, "Spoofing deep face recognition with custom silicone masks," in *2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS)*. IEEE, 2018, pp. 1–7. [Cited on pages vii, 19, 85, 97, 98, 103, 106, 107].

[52] A. George, Z. Mostaani, D. Geissenbuhler, O. Nikisins, A. Anjos, and S. Marcel, "Biometric face presentation attack detection with multi-channel convolutional neural network," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 42–55, 2020. [Cited on pages 19, 97, 98, 99, 100, 106].

[53] A. Agarwal, D. Yadav, N. Kohli, R. Singh, M. Vatsa, and A. Noore, "Face presentation attack with latex masks in multispectral videos," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2017, pp. 81–89. [Cited on pages 19, 97, 98, 99, 100, 106].

[54] S. Bhattacharjee and S. 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)*. IEEE, 2017, pp. 1–7. [Cited on pages 19, 97, 98, 106, 107].

[55] I. Pavlidis and P. 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)*. IEEE, 2000, pp. 15–24. [Cited on pages 19, 21, 97, 98].

[56] T. I. Dhamecha, A. Nigam, R. Singh, and M. Vatsa, "Disguise detection and face recognition in visible and thermal spectrums," in *2013 International Conference on Biometrics (ICB)*. IEEE, 2013, pp. 1–8. [Cited on pages vii, 19, 20, 96, 97, 98, 99].

[57] N. J. Short, A. J. Yuffa, G. Videen, and S. Hu, "Effects of surface materials on polarimetric-thermal measurements: applications to face recognition," *Applied optics*, vol. 55, no. 19, pp. 5226–5233, 2016. [Cited on pages vii, 20].

[58] V. Duro, "Face recognition by means of advanced contributions in machine learning," 2013. [Cited on pages vii, 21].

[59] F. J. Prokoski and R. B. Riedel, "Infrared identification of faces and body parts," in *Biometrics*. Springer, 1996, pp. 191–212. [Cited on pages 21].

[60] Face recognition databases. Http://www.face-rec.org/databases/. [Cited on pages 24].

[61] Human identification at a distance database. Http://www.equinoxsensors.com/products/HID.html. [Cited on pages 24, 25, 26].

[62] X. Chen, P. J. Flynn, and K. W. Bowyer, "Visible-light and infrared face recognition." ACM Workshop on Multimodal User Authentication, 2003, pp. 48–55. [Cited on pages ix, 24, 25, 78, 91, 92].

[63] P. J. Flynn, K. W. Bowyer, and P. J. Phillips, "Assessment of time dependency in face recognition: An initial study," in *International Conference on Audio-and Video-Based Biometric Person Authentication*. Springer, 2003, pp. 44–51. [Cited on pages ix, 24, 25, 91, 92].

[64] X. C. P. J. F. Kevin and W. Bowyer, "Visible-light and infrared face recognition," in *Workshop on Multimodal User Authentication*. Citeseer, 2003, p. 48. [Cited on pages ix, 24, 25, 91, 92].

[65] S. Wang, Z. Liu, S. Lv, Y. Lv, G. Wu, P. Peng, F. Chen, and X. Wang, "A natural visible and infrared facial expression database for expression recognition and emotion inference," *IEEE transactiond on Multimedia*, pp. 682–691, 2010. [Cited on pages 24, 25].

[66] B. Abidi, "Iris thermal/visible face database." DOE University Research Program in Robotics under grant DOE-DE-FG02-86NE37968, 2007. [Cited on pages 25].

[67] V. Espinosa-Duró, M. Faundez-Zanuy, and J. Mekyska, "A new face database simultaneously acquired in visible, near-infrared and thermal spectrums," *Cognitive Computation*, vol. 5, no. 1, pp. 119–135, 2013. [Cited on pages 25].

[68] V. Espinosa-Duró, M. Faundez-Zanuy, J. Mekyska, and E. Monte-Moreno, "A criterion for analysis of different sensor combinations with an application to face biometrics," *Cognitive Computation*, vol. 2, no. 3, pp. 135–141, 2010. [Cited on pages 25].

[69] Flir systems. Http://www.flir.com/. [Cited on pages 26].

[70] E. Mostafa, R. Hammoud, A. Ali, and A. Farag, "Face recognition in low resolution thermal images," *Journal Computer Vision and Image Understanding*, pp. 1689–1694, 2013. [Cited on pages 26].

[71] Commission nationale de l'informatique et des libertés. Https://www.cnil.fr/. [Cited on pages 27].

[72] M. A. Turk and A. P. Pentland, "Face recognition using eigenfaces," in *Proceedings. 1991 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1991, pp. 586–591. [Cited on pages 30].

[73] L. Sirovich and M. Kirby, "Low-dimensional procedure for the characterization of human faces," *Journal of the Optical Society in America. A, optics and image science*, 1987. [Cited on pages 30].

[74] K. Etemad and R. Chellappa, "Discriminant analysis for recognition of human face images." Int. Conference on Audio and Video-Based Biometric Person Authentication, 1997. [Cited on pages 30].

[75] H. Moon and P. J. Phillips, "Computational and performance aspects of pca-based face-recognition algorithms," *In Perception*, 2001. [Cited on pages 31].

[76] S. Yong, J. H. Lee, and J. B. Ra, "Multi-sensor image registration based on intensity and edge orientation information," *Pattern Recognition*, 2008. [Cited on pages 33].

[77] M. Kowalski, A. Grudzień, N. Palka, and M. Szustakowski, "Face recognition in the thermal infrared domain," in *Counterterrorism, Crime Fighting, Forensics, and Surveillance Technologies*, vol. 10441. International Society for Optics and Photonics, 2017, p. 1044109. [Cited on pages 38].

[78] G. H. Vigneau, J. L. Verdugo, G. F. Castro, F. Pizarro, and E. Vera, "Thermal face recognition under temporal variation conditions," *Ieee access*, vol. 5, pp. 9663–9672, 2017. [Cited on pages 38].

[79] L. B. Wolff, D. A. Socolinsky, and C. K. Eveland, "Quantitative measurement of illumination invariance for face recognition using thermal infrared imagery," in *Proceedings of SPIE*, vol. 4820, 2002, pp. 140–151. [Cited on pages 38].

[80] M. Sarfraz and R. Stiefelhagen, "Deep perceptual mapping for thermal to visible face recognition," *International Journal of Computer Vision*, pp. 1–11, 2015. [Cited on pages 38].

[81] J. Li, P. Hao, C. Zhang, and M. Dou, "Hallucinating faces from thermal infrared images," in *Proceedings of the International Conference on Image Processing, ICIP 2008, October 12-15, 2008, San Diego, California, USA*. IEEE, 2008, pp. 465–468. [Online]. Available: https://doi.org/10.1109/ICIP.2008.4711792 [Cited on pages 38, 39].

[82] S. Y. L. D. J. Choi, S. Hu, "Thermal to visible face recognition." Proceedings of SPIE - The International Society for Optical Engineering, 2012, pp. 311–316. [Cited on pages 38].

[83] C. Chen and A. Ross, "Matching thermal to visible face images using hidden factor analysis in a cascaded subspace learning framework," *Pattern Recognition Letters*, 2016. [Cited on pages 38].

[84] T. Zhang, A. Wiliem, S. Yang, and B. Lovell, "TV-GAN: Generative adversarial network based thermal to visible face recognition," in *2018 International Conference on Biometrics (ICB)*, Feb 2018, pp. 174–181. [Cited on pages viii, 38, 39, 40, 45, 46, 47, 49, 50, 51, 52, 53, 54, 57, 60, 114].

[85] A. Kantarcı and H. K. Ekenel, "Thermal to visible face recognition using deep autoencoders," in *2019 International Conference of the Biometrics Special Interest Group (BIOSIG)*. IEEE, 2019, pp. 1–5. [Cited on pages 38].

[86] M. Dou, C. Zhang, P. Hao, and J. Li, "Converting thermal infrared face images into normal gray-level images." Asian Conference on Computer Vision, 2007, pp. 722–732. [Cited on pages 39].

[87] H. Zhang, V. M. Patel, B. S. Riggan, and S. Hu, "Generative adversarial network-based synthesis of visible faces from polarimetric thermal faces," in *2017 IEEE International Joint Conference on Biometrics (IJCB)*, Oct 2017, pp. 100–107. [Cited on pages 39].

[88] Z. Wang, Z. Chen, and F. Wu, "Thermal to visible facial image translation using generative adversarial networks," *IEEE Signal Processing Letters*, vol. 25, no. 8, pp. 1161–1165, 2018. [Cited on pages 39, 40].

[89] L. Song, M. Zhang, X. Wu, and R. He, "Adversarial discriminative heterogeneous face recognition." AAAI Conference on Artificial Intelligence, 2018. [Cited on pages 39].

[90] A. C. Guei and M. A. Akhloufi, "Deep generative adversarial networks for infrared image enhancement," *Proc. SPIE*, vol. 10661, pp. 10 661 – 10 661 – 12, 2018. [Cited on pages 39].

[91] H. Zhang, B. S. Riggan, S. Hu, N. J. Short, and V. M. Patel, "Synthesis of high-quality visible faces from polarimetric thermal faces using generative adversarial networks," *International Journal of Computer Vision*, 2018. [Cited on pages 39].

[92] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *NIPS*, 2014, p. 2672–2680. [Cited on pages 39, 45].

[93] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," *arXiv preprint arXiv:1511.06434*, 2015. [Cited on pages 40].

[94] D. Berthelot, T. Schumm, and L. Metz, "Began: boundary equilibrium generative adversarial networks," *European Conference of Computer Vision Workshops (ECCVW)*, 2018. [Cited on pages 40].

[95] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networkss," in *Computer Vision (ICCV), 2017 IEEE International Conference on*, 2017. [Cited on pages 40].

[96] P. Isola, J. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017*. IEEE Computer Society, 2017, pp. 5967–5976. [Online]. Available: https://doi.org/10.1109/CVPR.2017.632 [Cited on pages viii, 40, 42, 45, 46, 47, 49, 50, 51, 52, 54, 57, 60, 114].

[97] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," 2015. [Cited on pages 40].

[98] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 815–823. [Cited on pages 40].

[99] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *European conference on computer vision*. Springer, 2016, pp. 694–711. [Cited on pages 40].

[100] R. Mechrez, I. Talmi, and L. 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*, ser. Lecture Notes in Computer Science, vol. 11218. Springer, 2018, pp. 800–815. [Online]. Available: https://doi.org/10.1007/978-3-030-01264-9_47 [Cited on pages 40, 42, 44, 99, 114].

[101] L. Xu, J. S. J. Ren, C. Liu, and J. Jia, "Deep convolutional neural network for image deconvolution," in *Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, December 8-13 2014, Montreal, Quebec, Canada*, 2014, pp. 1790–1798. [Online]. Available: http://papers.nips.cc/paper/5485-deep-convolutional-neural-network-for-image-deconvolution [Cited on pages 42].

[102] L. A. Gatys, A. S. Ecker, and M. 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. IEEE Computer Society, 2016, pp. 2414–2423. [Online]. Available: https://doi.org/10.1109/CVPR.2016.265 [Cited on pages 42, 43, 44].

[103] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *CoRR*, vol. abs/1409.1556, 2014. [Online]. Available: http://arxiv.org/abs/1409.1556 [Cited on pages 44].

[104] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A Large-Scale Hierarchical Image Database," in *CVPR09*, 2009. [Cited on pages 44].

[105] K. Mallat and J.-L. 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)*. IEEE, 2018, pp. 1–5. [Cited on pages ix, 44, 68, 72, 73, 78, 80, 91].

[106] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds. Cham: Springer International Publishing, 2015, pp. 234–241. [Cited on pages 45].

[107] L. Tran, X. Yin, and X. Liu, "Disentangled representation learning gan for pose-invariant face recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1415–1424. [Cited on pages 45].

[108] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004. [Cited on pages 50].

[109] B. Amos, B. Ludwiczuk, and M. Satyanarayanan, "Openface: A general-purpose face recognition library with mobile applications," CMU-CS-16-118, CMU School of Computer Science, Tech. Rep., 2016. [Cited on pages 51, 62].

[110] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," in *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2015, Boston, MA, USA, June 7-12, 2015*, 2015, pp. 815–823. [Online]. Available: https://doi.org/10.1109/CVPR.2015.7298682 [Cited on pages 51].

[111] D. Yi, Z. Lei, S. Liao, and S. Z. Li, "Learning face representation from scratch," *CoRR*, vol. abs/1411.7923, 2014. [Online]. Available: http://arxiv.org/abs/1411.7923 [Cited on pages 51].

[112] H. Ng and S. Winkler, "A data-driven approach to cleaning large face datasets," in *2014 IEEE International Conference on Image Processing, ICIP 2014, Paris, France, October 27-30, 2014*, 2014, pp. 343–347. [Online]. Available: https://doi.org/10.1109/ICIP.2014.7025068 [Cited on pages 51].

[113] X. Wu, R. He, Z. Sun, and T. Tan, "A light cnn for deep face representation with noisy labels," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 11, pp. 2884–2896, 2018. [Cited on pages 52, 59, 62, 115].

[114] K. Mallat, N. Damer, F. Boutros, A. Kuijper, and J.-L. Dugelay, "Cross-spectrum thermal to visible face recognition based on cascaded image synthesis," in *2019 International Conference on Biometrics (ICB)*. IEEE, 2019, pp. 1–8. [Cited on pages 59, 60, 62, 80].

[115] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern recognition*, vol. 29, no. 1, pp. 51–59, 1996. [Cited on pages 59, 64, 115].

[116] N. Damer, F. Boutros, K. Mallat, F. Kirchbuchner, J.-L. Dugelay, and A. Kuijper, "Cascaded generation of high-quality color visible face images from thermal captures," *arXiv preprint arXiv:1910.09524*, 2019. [Cited on pages 60].

[117] S. M. Desa and S. Hati, "Ir and visible face recognition using fusion of kernel based features," in *2@inproceedingsgyaourova2004fusion, title=Fusion of infrared and visible images for face recognition, author=Gyaourova, Aglika and Bebis, George and Pavlidis, Ioannis, booktitle=European Conference on Computer Vision, pages=456–468, year=2004, organization=Springer 008 19th International Conference on Pattern Recognition*. IEEE, 2008, pp. 1–4. [Cited on pages 60].

[118] X. Chen, Z. Jing, and G. Xiao, "Fuzzy fusion for face recognition," in *International Conference on Fuzzy Systems and Knowledge Discovery*. Springer, 2005, pp. 672–675. [Cited on pages 61].

[119] P. Buyssens, M. Revenu, and O. Lepetit, "Fusion of ir and visible light modalities for face recognition," in *2009 IEEE 3rd International Conference on Biometrics: Theory, Applications, and Systems*. IEEE, 2009, pp. 1–6. [Cited on pages 61].

[120] H. Schwenk, "The diabolo classifier," *Neural Computation*, vol. 10, no. 8, pp. 2175–2200, 1998. [Cited on pages 61].

[121] O. Arandjelovic and R. Hammoud, "Multi-sensory face biometric fusion (for personal identification)," in *2006 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'06)*. IEEE, 2006, pp. 128–128. [Cited on pages 61].

[122] D. A. Socolinsky and A. Selinger, "A comparative analysis of face recognition performance with visible and thermal infrared imagery," in *Object recognition supported by user interaction for service robots*, vol. 4. IEEE, 2002, pp. 217–222. [Cited on pages 61].

[123] D. A. Socolinsky, A. Selinger, and J. D. Neuheisel, "Face recognition with visible and thermal infrared imagery," *Computer vision and image understanding*, vol. 91, no. 1-2, pp. 72–114, 2003. [Cited on pages 61].

[124] J. Fierrez-Aguilar, J. Ortega-Garcia, J. Gonzalez-Rodriguez, and J. Bigun, "Discriminative multimodal biometric authentication based on quality measures," *Pattern recognition*, vol. 38, no. 5, pp. 777–779, 2005. [Cited on pages 61].

[125] M. Vatsa, R. Singh, and A. Noore, "Integrating image quality in 2ν-svm biometric match score fusion," *International Journal of Neural Systems*, vol. 17, no. 05, pp. 343–351, 2007. [Cited on pages 61].

[126] Z. Zhou, E. Y. Du, C. Belcher, N. L. Thomas, and E. J. Delp, "Quality fusion based multimodal eye recognition," in *2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 2012, pp. 1297–1302. [Cited on pages 61].

[127] P. Wasnik, K. B. Raja, R. Ramachandra, and C. Busch, "Assessing face image quality for smartphone based face recognition system," in *Biometrics and Forensics (IWBF), 2017 5th International Workshop on*. IEEE, 2017, pp. 1–6. [Cited on pages 64].

[128] F. T. Zohra, A. D. Gavrilov, O. Z. Duran, and M. Gavrilova, "A linear regression model for estimating facial image quality," in *Cognitive Informatics & Cognitive Computing (ICCI* CC), 2017 IEEE 16th International Conference on*. IEEE, 2017, pp. 130–138. [Cited on pages 64].

[129] X. Gao, S. Z. Li, R. Liu, and P. Zhang, "Standardization of face image sample quality," in *International Conference on Biometrics*. Springer, 2007, pp. 242–251. [Cited on pages 64].

[130] K. Matkovic, L. Neumann, A. Neumann, T. Psik, and W. Purgathofer, "Global contrast factor-a new approach to image contrast." *Computational Aesthetics*, vol. 2005, pp. 159–168, 2005. [Cited on pages 64].

[131] M. V. Shirvaikar, "An optimal measure for camera focus and exposure," in *Thirty-Sixth Southeastern Symposium on System Theory, 2004. Proceedings of the*. IEEE, 2004, pp. 472–475. [Cited on pages 64].

[132] N. D. Narvekar and L. J. Karam, "A no-reference image blur metric based on the cumulative probability of blur detection (cpbd)," *IEEE Transactions on Image Processing*, vol. 20, no. 9, pp. 2678–2683, 2011. [Cited on pages 64].

[133] S. Setiowati, E. L. Franita, I. Ardiyanto, *et al.*, "A review of optimization method in face recognition: Comparison deep learning and non-deep learning methods," in *2017 9th International Conference on Information Technology and Electrical Engineering (ICITEE)*. IEEE, 2017, pp. 1–6. [Cited on pages 70].

[134] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active appearance models," vol. 23, no. 6. IEEE, 2001, pp. 681–685. [Cited on pages 75, 76, 77, 81, 115].

[135] M. Kowalski, J. Naruniec, and T. Trzcinski, "Deep alignment network: A convolutional neural network for robust face alignment," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2017, pp. 88–97. [Cited on pages 75, 76, 77, 82, 115].

[136] F. Liu, Q. Zhao, D. Zeng, *et al.*, "Joint face alignment and 3d face reconstruction with application to face recognition," *IEEE transactions on pattern analysis and machine intelligence*, 2018. [Cited on pages 76].

[137] X. Yin, X. Yu, K. Sohn, X. Liu, and M. Chandraker, "Towards large-pose face frontalization in the wild," pp. 3990–3999, 2017. [Cited on pages 76].

[138] P. Liu, Y. Yu, Y. Zhou, and S. Du, "Single view 3d face reconstruction with landmark updating," in *2019 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*. IEEE, 2019, pp. 403–408. [Cited on pages 76].

[139] M. Munasinghe, "Facial expression recognition using facial landmarks and random forest classifier," in *2018 IEEE/ACIS 17th International Conference on Computer and Information Science (ICIS)*. IEEE, 2018, pp. 423–427. [Cited on pages 76].

[140] J. S. Chung, A. Senior, O. Vinyals, and A. Zisserman, "Lip reading sentences in the wild," pp. 3444–3453, 2017. [Cited on pages 76].

[141] P. N. Belhumeur, D. W. Jacobs, D. J. Kriegman, and N. Kumar, "Localizing parts of faces using a consensus of exemplars," vol. 35, no. 12. IEEE, 2013, pp. 2930–2940. [Cited on pages 76, 78, 79].

[142] V. Le, J. Brandt, Z. Lin, L. Bourdev, and T. S. Huang, "Interactive facial feature localization," in *European conference on computer vision*. Springer, 2012, pp. 679–692. [Cited on pages 76, 78, 79].

[143] D. Cristinacce and T. F. Cootes, "Feature detection and tracking with constrained local models." in *Bmvc*, vol. 1, no. 2. Citeseer, 2006, p. 3. [Cited on pages 77].

[144] X. Cao, Y. Wei, F. Wen, and J. Sun, "Face alignment by explicit shape regression," *International Journal of Computer Vision*, 2014. [Cited on pages 77].

[145] X. Xiong and F. De la Torre, "Supervised descent method and its applications to face alignment," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2013, pp. 532–539. [Cited on pages 77].

[146] X. Dong, Y. Yan, W. Ouyang, and Y. Yang, "Style aggregated network for facial landmark detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 379–388. [Cited on pages 77].

[147] Y. Wu and Q. Ji, "Facial landmark detection: A literature survey," *International Journal of Computer Vision*, vol. 127, no. 2, pp. 115–142, 2019. [Cited on pages 77].

[148] H.-W. Tzeng, H.-C. Lee, and M.-Y. Chen, "The design of isotherm face recognition technique based on nostril localization," in *Proceedings 2011 International Conference on System Science and Engineering*. IEEE, 2011, pp. 82–86. [Cited on pages 77].

[149] S. Wang, Z. Liu, P. Shen, and Q. Ji, "Eye localization from thermal infrared images," *Pattern Recognition*, vol. 46, no. 10, pp. 2613–2621, 2013. [Cited on pages 77].

[150] A. H. Alkali, R. Saatchi, H. Elphick, and D. Burke, "Eyes' corners detection in infrared images for real-time noncontact respiration rate monitoring," in *2014 World Congress on Computer Applications and Information Systems (WCCAIS)*. IEEE, 2014, pp. 1–5. [Cited on pages 77].

[151] M. Kopaczka, K. Acar, and D. Merhof, "Robust facial landmark detection and face tracking in thermal infrared images using active appearance models." in *VISIGRAPP (4: VISAPP)*, 2016, pp. 150–158. [Cited on pages 77, 78, 82, 83, 87, 92].

[152] M. Kopaczka, L. Breuer, J. Schock, and D. Merhof, "A modular system for detection, tracking and analysis of human faces in thermal infrared recordings," *Sensors*, vol. 19, no. 19, p. 4135, 2019. [Cited on pages 77].

[153] M. Kopaczka, R. Kolk, and D. Merhof, "A fully annotated thermal face database and its application for thermal facial expression recognition," in *2018 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*. IEEE, 2018, pp. 1–6. [Cited on pages 77, 78, 83, 87, 88, 90].

[154] M. Kowalski and A. Grudzień, "High-resolution thermal face dataset for face and expression recognition," *Metrology and Measurement Systems*, vol. 25, no. 2, 2018. [Cited on pages ix, 78, 91, 92].

[155] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker, "Multi-pie," *Image and Vision Computing*, vol. 28, no. 5, pp. 807–813, 2010. [Cited on pages 79].

[156] K. Messer, J. Matas, J. Kittler, J. Luettin, and G. Maitre, "Xm2vtsdb: The extended m2vts database," in *Second international conference on audio and video-based biometric person authentication*, vol. 964, 1999, pp. 965–966. [Cited on pages 79].

[157] M. Koestinger, P. Wohlhart, P. M. Roth, and H. Bischof, "Annotated facial landmarks in the wild: A large-scale, real-world database for facial landmark localization," in *2011 IEEE international conference on computer vision workshops (ICCV workshops)*. IEEE, 2011, pp. 2144–2151. [Cited on pages 79].

[158] C. Sagonas, G. Tzimiropoulos, S. Zafeiriou, and M. Pantic, "300 faces in-the-wild challenge: The first facial landmark localization challenge," in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2013, pp. 397–403. [Cited on pages ix, 79, 80, 83].

[159] J. C. Gower, "Generalized procrustes analysis," *Psychometrika*, vol. 40, no. 1, pp. 33–51, 1975. [Cited on pages 82].

[160] S. Baker, *Inverse Compositional Algorithm*. Boston, MA: Springer US, 2014, pp. 426–428. [Online]. Available: https://doi.org/10.1007/978-0-387-31439-6_759 [Cited on pages 82].

[161] B. Johnston and P. de Chazal, "A review of image-based automatic facial landmark identification techniques," vol. 2018, no. 1. Springer, 2018, p. 86. [Cited on pages 83].

[162] G. Trigeorgis, P. Snape, M. A. Nicolaou, E. Antonakos, and S. Zafeiriou, "Mnemonic descent method: A recurrent process applied for end-to-end face alignment," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 4177–4187. [Cited on pages 83].

[163] S. Ren, X. Cao, Y. Wei, and J. Sun, "Face alignment at 3000 fps via regressing local binary features," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1685–1692. [Cited on pages 83].

[164] S. Zafeiriou, G. Trigeorgis, G. Chrysos, J. Deng, and J. Shen, "The menpo facial landmark localisation challenge: A step towards the solution," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2017, pp. 170–179. [Cited on pages 83].

[165] S. Zhu, C. Li, C. Change Loy, and X. Tang, "Face alignment by coarse-to-fine shape searching," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 4998–5006. [Cited on pages 83].

[166] D. E. King, "Dlib toolkit," https://github.com/davisking/dlib. [Cited on pages ix, 85].

[167] A. K. Jain, K. Nandakumar, and A. Nagar, "Biometric template security," *EURASIP Journal on advances in signal processing*, vol. 2008, pp. 1–17, 2008. [Cited on pages 96].

[168] S. Marcel, M. S. Nixon, and S. Z. Li, *Handbook of biometric anti-spoofing*. Springer, 2014, vol. 1. [Cited on pages 96].

[169] A. Mohammadi, S. Bhattacharjee, and S. Marcel, "Deeply vulnerable: a study of the robustness of face recognition to presentation attacks," *IET Biometrics*, vol. 7, no. 1, pp. 15–26, 2017. [Cited on pages 96].

[170] N. K. Ratha, J. H. Connell, and R. M. Bolle, "An analysis of minutiae matching strength," in *International Conference on Audio-and Video-Based Biometric Person Authentication*. Springer, 2001, pp. 223–228. [Cited on pages 96].

[171] N. Erdogmus, N. Kose, and J.-L. Dugelay, "Impact analysis of nose alterations on 2d and 3d face recognition," in *2012 IEEE 14th International Workshop on Multimedia Signal Processing (MMSP)*. IEEE, 2012, pp. 354–359. [Cited on pages 96].

[172] V. Chiesa and J.-L. Dugelay, "Advanced face presentation attack detection on light field database," in *2018 International Conference of the Biometrics Special Interest Group (BIOSIG)*. IEEE, 2018, pp. 1–4. [Cited on pages 97].

[173] S. Kim, Y. Ban, and S. Lee, "Face liveness detection using a light field camera," *Sensors*, vol. 14, no. 12, pp. 22 471–22 499, 2014. [Cited on pages 97].

[174] P. Buddharaju, I. T. Pavlidis, P. Tsiamyrtzis, and M. Bazakos, "Physiology-based face recognition in the thermal infrared spectrum," *IEEE transactions on pattern analysis and machine intelligence*, vol. 29, no. 4, pp. 613–626, 2007. [Cited on pages 116].

[175] S.-Q. Wu, L.-Z. Wei, Z.-J. Fang, R.-W. Li, and X.-Q. Ye, "Infrared face recognition based on blood perfusion and sub-block dct in wavelet domain," in *2007 International Conference on Wavelet Analysis and Pattern Recognition*, vol. 3. IEEE, 2007, pp. 1252–1256. [Cited on pages 116].