# A benchmark database of visible and thermal paired face images across multiple variations

Khawla Mallat[1], Jean-Luc Dugelay[1]

**Abstract:** Although visible face recognition systems have grown as a major area of research, they are still facing serious challenges when operating in uncontrolled environments. In attempt to overcome these limitations, thermal imagery has been investigated as a promising direction to extend face recognition technology. However, the reduced number of databases acquired in thermal spectrum limits its exploration. In this paper, we introduce a database of face images acquired simultaneously in visible and thermal spectra under various variations: illumination, expression, pose and occlusion. Then, we present a comparative study of face recognition performances on both modalities against each variation and the impact of bimodal fusion. We prove that thermal spectrum rivals with the visible spectrum not only in the presence of illumination changes, but also in case of expression and poses changes.

**Keywords:** thermal spectrum, visible spectrum, face database, illumination, expression, pose, occlusion, sensor-level fusion, score-level fusion.

## 1 Introduction

Biometric systems have been widely used for law enforcement and security systems over the past few years. This technology is the science of analyzing physical or behavioral characteristics specific to each individual in order to be able to authenticate their identity. One of the most used trait for biometric systems is face. Therefore, in the last two decades, automatic face recognition has consistently been one of the most active research areas of computer vision. Systems based on images acquired in the visible spectrum have reached a significant level of maturity with some practical success. However, a range of factors continue to issue serious problems when visible spectrum based face recognition methods are applied in unconstrained settings. Dealing with head pose variation, facial expression changes, occlusion and illumination changes is still challenging.

Different lines of research attempted to overcome these aforesaid challenges, by developing advanced techniques applied on visible face images. In practice, pose and expression changes are usually less problematic and can often be overcome by acquiring data over a time period by tracking a face. However illumination changes are far more difficult to tackle, some works suggested the use of homomorphic filtering, Gamma Intensity Correction and histogram equalization [ZKM07]. Another promising direction to explore is thermal imagery, considering its high robustness to illumination changes.

Thermal imagery detects electromagnetic radiations in the medium wave MWIR (3 - 8$\mu$m) and long wave infrared spectrum LWIR (8 - 15$\mu$m) in which most of the heat energy is

[1] Department of Digital Security, EURECOM, Sophia-Antipolis, France, {mallat, dugelay}@eurecom.fr

emitted by the skin [RMY17]. Therefore, it is possible to acquire a face representation without any external source of illumination. However, the thermal heat emitted by the face can be affected by various factors such as ambient temperature, physical exercise or illness. This sensitivity makes thermal face recognition over time a challenging task.

In this paper, we provide, in section 2, an overview of public databases acquired simultaneously in visible and thermal spectra used to combine both modalities in order to improve face recognition performance. Then, we introduce our new database that addresses the lack of variability in the existing ones aiming to develop face recognition systems robust against real-world challenges. Section 4 describes the experiments conducted to provide a comparative study of face recognition performance under various variations introduced in the database and presents the performance of each modality and the impact of bimodal fusion. The final section concludes the paper and points out directions for future work.

## 2 Overview of the existing visible and thermal face databases

Exploiting the advantages from both visible and thermal spectra, and seeking to limit the drawbacks of each, [BR10, He15] suggested systems based on fusing visible and thermal modalities in different levels and reached improvements in face recognition performance. However, most of the existing biometric systems are based on databases acquired in visible spectrum. Therefore, particular studies has thoroughly focused on cross-spectrum face recognition algorithms. This discipline aims to identify a person imaged in thermal spectrum from a gallery containing face images acquired in the visible spectrum. Considering the large difference between the two modalities, cross-spectrum face recognition aimed to overcome this gap by artificially estimating equivalent visible face images based on their thermal counterparts [Do07, Li07]. These synthesized face images are then compared to the gallery of visible images. Another approach was based on learning projections between thermal and visible face images to bring the two modalities closer together [Ch12, Hu15, SS15, CR16]

To assess the effectiveness of face recognition systems, an evaluation on an input data, presented in different conditions, is required. Currently, there are numerous public face databases acquired in visible spectrum covering all variations possible [Fa]. However, interest in utilizing thermal face images has grown recently and thus only a few databases have been provided, particularly databases that involve simultaneously acquired images in thermal and visible spectra. We present, in table 1, the key descriptors of the few public databases containing visible face images and their thermal counterpart.

| Name | Thermal resolution | # of images / # Subjects | # of variations | | | | |
|---|---|---|---|---|---|---|---|
| | | | Illumination | Expression | Pose | Occlusions | Time-lapse |
| NVIE [Wa10] | 320 × 240 | 7460 / 103 | 1 | 2×6 | 1 | 0 | 1 |
| UND-X1 [CFB03] | 312 × 239 | 4584 / 82 | 2 | 2 | 1 | 0 | Multiple |
| EQUINOX [Hu] | 320 × 240 | 25000 / 91 | 3 | 3 | 1 | 1 | 1 |

Tab. 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. NVIE database [Wa10] was acquired mainly to investigate the impact of thermal spectrum on expression recognition, thus the only variation considered was facial expression. Whereas UND-X1 database [CFB03] was focusing 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. The EQUINOX database [Hu] was collected in a single session, taking into account 3 expression variations and 3 light conditions. NVIE and UND-X1 databases were collected using different devices to acquire face images in visible and thermal spectra separately which does not guarantee having the same face image. Whilst EQUINOX database [Hu] was collected using a sensor capable of capturing simultaneous videos in both domains. Although head pose changes and occlusions are still a challenging factor for face recognition algorithms, none of the databases have considered these variations. In addition, whilst thermal face images are very sensitive to time-lapse factor, only UND-X1 database acquired data in multiple sessions.

## 3 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 these publicly available databases and the lack of variations considered. In this section, we present the acquisition setup of our database followed by a description of the collection protocol.

### 3.1 The acquisition setup

The presented database was acquired with a newly developed dual sensor, visible and thermal, camera FLIR Duo R by FLIR Systems [Fl]. This camera is designed for unmanned aerial vehicle allowing to capture simultaneously images and videos in both visible and thermal spectra. Accordingly, it is well suited for our database collection aimed at fusion for face recognition as well as cross-spectrum applications. The visible sensor is a CCD sensor with a pixel resolution of $1920\times1080$. The thermal sensor of this camera is an uncooled VoX microbolometer and has a spectral response range of 7.5 - 13.5$\mu$m with a pixel resolution of $160\times120$. Even though the thermal resolution of the camera may seem low, Mostafa et al. [Mo13] demonstrated that high face recognition rates can be achieved even with low resolution $64\times64$ pixels thermal face images.


Fig. 1: Flir Duo R camera and the acquisition setup.

The camera was set in front of the subject at the distance of 1.5 meter and placed at a height of 1 meter from the ground. The data collection took place in a controlled environment, where the ambient temperature was set in average to 25°C. The scene, as shown in Fig. 1, was equipped with three-point lighting kit: rim light, key light and fill light.

### 3.2 The database collection protocol

The subset of the database that we have provided at the present time was collected from 50 subjects of different ages, sex and ethnicities. Each subject participated in two different acquisition sessions separated by a time interval of 3 to 4 months. The data includes 21 face images per subject in each session with different facial variations, resulting a total of 4200 images. The camera was designed to take simultaneous face images of thermal and visible spectra. During the data collection, the camera was set to capture a shot every second to avoid acquisition errors. The variations considered during this session are shown in Fig. 2 and described as follow:

- 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.
- 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.

## 4 Preliminary evaluation of the database

The aim of this section is to present a preliminary evaluation to assess the applicability of the database. A comparison of thermal and visible domain against various facial variations introduced in our database is also performed. This comparison will reflect the performance of visible and thermal modalities and bimodal fusion in hands-on scenarios.

### 4.1 Face recognition algorithm

Face recognition algorithms can be split into two categories: feature based algorithms and holistic algorithms. To evaluate our database, we have settled for holistic algorithms in particular, Fisherfaces method followed by 1-Nearest Neighborhood classification. This choice was motivated by the fact that these algorithms, unlike feature based algorithms, do

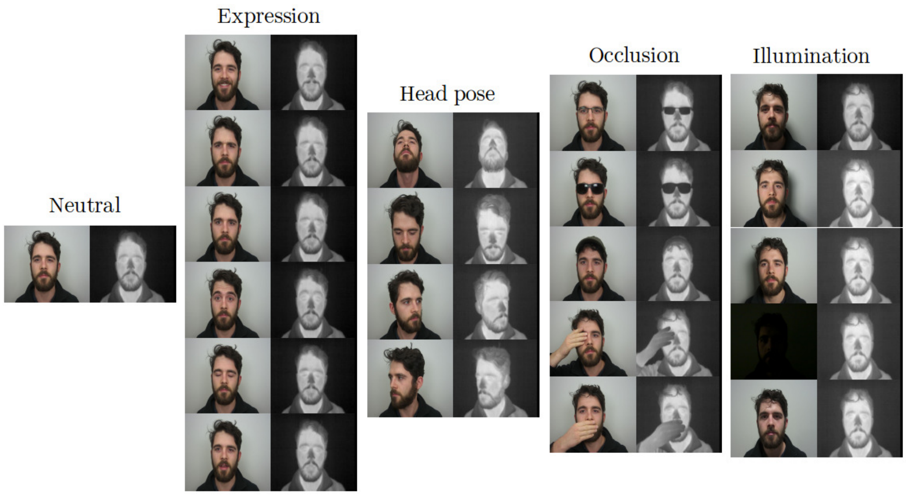


Fig. 2: Illustration of visible and thermal images for various facial variations.

not depend on the detection of facial landmarks, which can be challenging particularly for thermal images.

Fisherfaces [EC97] is based on both principal component analysis (PCA) and linear discriminant analysis (LDA). PCA linearly projects the image space onto a low dimensional feature space. Then, LDA performs dimensionality reduction while preserving the class discriminatory information by finding direction along which the classes are best separated. Fisherfaces algorithm has recorded high performances on visible face images. Moreover, Socolinsky et al. [SS02] have compared holistic face recognition algorithms and proved that Fisherfaces achieved the highest recognition rate on thermal face images.

### 4.2 Experiment and results

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 dataset was split in 4 subsets, 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.

Aiming to get a deeper understanding on how performances of both visible and thermal spectra differ face to different variations, we proposed to apply sensor-level fusion and score-level fusion since it combines the information provided by both modalities. Fusing visible and thermal modalities was based on simple averaging in both settings, so as to not boost one modality over the other. For sensor-level fusion, visible and thermal face images were co-registered using edge-based image registration approach inspired from [YLR08] and then we used Discrete Wavelet Transform (DWT) based image fusion, as described in [Kl12], to generate fused images illustrated in Fig. 3. While for score-level fusion, the matching scores of each of the two modalities test images are first normalized and then merged to obtain an average matching score.

(a) (b) (c)

Fig. 3: a: visible image, b: thermal image, c: sensor-level fusion resulting image.

The experiments described above were performed on data acquired in the first session. Table 2 and table 3 illustrate the Rank-1 recognition rates of Fisherfaces algorithm on each modality as well as on bimodal fusion.

| | | TEST | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Illumination | | | | Expression | | | |
| | | VIS | TH | Sensor Fusion | Score Fusion | VIS | TH | Sensor Fusion | Score Fusion |
| T | Illumination | – | – | – | – | 0.814 | 0.96 | 0.9 | 0.96 |
| R | Expression | 0.733 | 0.973 | 0.83 | 0.94 | – | – | – | – |
| A | Pose | 0.66 | 0.893 | 0.823 | 0.883 | 0.914 | 0.891 | 0.934 | 0.957 |
| I | Occlusion | 0.793 | 0.973 | 0.86 | 0.943 | 0.957 | 0.962 | 0.965 | 0.982 |
| N | **Average** | **0.728** | **0.946** | **0.837** | **0.922** | **0.895** | **0.937** | **0.934** | **0.966** |

Tab. 2: Rank-1 recognition under illumination and expression variations.

| | | TEST | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Pose | | | | Occlusion | | | |
| | | VIS | TH | Sensor Fusion | Score Fusion | VIS | TH | Sensor Fusion | Score Fusion |
| T | Illumination | 0.312 | 0.365 | 0.405 | 0.384 | 0.69 | 0.59 | 0.656 | 0.703 |
| R | Expression | 0.476 | 0.417 | 0.554 | 0.56 | 0.83 | 0.53 | 0.71 | 0.873 |
| A | Pose | – | – | – | – | 0.633 | 0.42 | 0.546 | 0.686 |
| I | Occlusion | 0.38 | 0.365 | 0.489 | 0.428 | – | – | – | – |
| N | **Average** | **0.389** | **0.382** | **0.482** | **0.457** | **0.719** | **0.513** | **0.637** | **0.754** |

Tab. 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 external source of lighting to acquire images while visible spectrum is highly sensitive to light changes. Likewise 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 fact that visible spectrum is highly sensitive to light changes while the thermal spectrum is not, since changes in facial expressions or head pose implies changes in the distribution of the light across the face surface. Although, when it comes for head pose variation, we notice that both visible and thermal spectrum perform almost equally. 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 there is no heat emitted, while on visible spectrum we can see the eye details thanks visible light transmittance in glass.

Focusing on fusion results, when a modality performs considerably higher than the other, we did not obtain any improvement in performance of face recognition when applying fusion. Nevertheless, when the performance of the two modalities is notably similar, applying fusion provides improved results. 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.

## 5 Conclusion and Future work

We have proposed a new database of face images acquired simultaneously in thermal and visible domains aiming to cover a wider range of facial variations compliant with hands-on scenarios. A comparative study of face recognition performance of visible spectrum, thermal spectrum, sensor-level fusion and score-level fusion under various variations is also presented in this paper. Although, the utility of thermal spectrum is commonly acknowledged in the presence of illumination constraints, we have proved that thermal spectrum outperforms visible spectrum not only under light variation but also in expression changes, and proceed correspondingly when it comes for head pose changes. Therefore, thermal imagery can be considered not only as a complement for visible imagery but also an alternative.

Future work will be involving a study of time-lapse impact on thermal spectrum performance using data from the two sessions. Also, we will provide shortly a high resolution version of the database using an updated camera system that has been recently released.

## 6 Acknowledgment

Our research activities are partly supported by the collaborative projects FUI COOPOL, EIT Drones112 and H2020 PROTECT.

## References


[BR10] Buyssens, P.; Revenu, M.: Fusion levels of visible and infrared modalities for face recognition. 2010.

[CFB03] Chen, X.; Flynn, P. J.; Bowyer, K. W.: Visible-light and Infrared Face Recognition. ACM Workshop on Multimodal User Authentication, pp. 48–55, 2003.

[Ch12] Choi, J.; Hu, S.; Young, S.; Davis, L.: Thermal to visible face recognition. Proceedings of SPIE - The International Society for Optical Engineering, pp. 311–316, 2012.

[CR16] Chen, C.; Ross, A.: Matching thermal to visible face images using hidden factor analysis in a cascaded subspace learning framework. Pattern Recognition Letters, 2016.

[Do07] Dou, M.; Zhang, C.; Hao, P.; Li, J.: Converting Thermal Infrared Face Images into Normal Gray-Level Images. Asian Conference on Computer Vision, pp. 722–732, 2007.

[EC97] Etemad, K.; Chellappa, R.: Discriminant analysis for recognition of human face images. Int. Conference on Audio and Video-Based Biometric Person Authentication, 1997.

[Fa] Face recognition databases, http://www.face-rec.org/databases/.

[Fl] Flir systems, http://www.flir.com/.

[He15] Hermosilla, G.; Gallardo, F.; Farias, G.; Martin, C. San: Fusion of Visible and Thermal Descriptors Using Genetic Algorithms for Face Recognition Systems. Sensors, 2015.

[Hu] Human identification at a distance database, http://www.equinoxsensors.com/products/HID.html.

[Hu15] Hu, S.; Choi, J.; Chan, A.; Schwartz, W.: Thermal-to-visible face recognition using partial least squares. Journal of the Optical Society of America, pp. 431–442, 2015.

[Kl12] Klein, Lawrence A., ed. Sensor and Data Fusion: A Tool for Information Assessment and Decision Making, Second Edition. SPIE, 2012.

[Li07] Li, J.; Hao, P.; Zhang, C.; Dou, M.: Hallucinating faces from thermal infrared images. Asian Conference on Computer Vision, pp. 722–732, 2007.

[Mo13] Mostafa, E.; Hammoud, R.; Ali, A.; Farag, A.: Face recognition in low resolution thermal images. Journal Computer Vision and Image Understanding, pp. 1689–1694, 2013.

[RMY17] Rai, M.; Maity, T.; Yadav, R. K.: Thermal imaging system and its real time applications: a survey. Journal of Engineering Technology, pp. 290–303, 2017.

[SS02] Socolinsky, D. A.; Selinger, A.: A comparative analysis of face recognition performance with visible and thermal infrared imagery. Proc. International Conference on Pattern Recognition, 2002.

[SS15] Sarfraz, M.; Stiefelhagen, R.: Deep Perceptual Mapping for Thermal to Visible Face Recognition. International Journal of Computer Vision, pp. 1–11, 2015.

[Wa10] Wang, S.; Liu, Z.; Lv, S.; Lv, Y.; Wu, G.; Peng, P.; Chen, F.; Wang, X.: A Natural Visible and Infrared Facial Expression Database for Expression Recognition and Emotion Inference. IEEE transactiond on Multimedia, pp. 682–691, 2010.

[YLR08] Yong, S.K.; Lee, J. H.; Ra, J. B.: Multi-sensor image registration based on intensity and edge orientation information. Pattern Recognition, 2008.

[ZKM07] Zou, X.; Kittler, J.; Messer, K.: Illumination Invariant Face Recognition: A Survey. IEEE International Conference on Biometrics: Theory, Applications, and Systems, 2007.