



**HAL**  
open science

## Beyond the visible: thermal data for facial soft biometric estimation

Nelida Mirabet-Herranz, Jean-Luc Dugelay

► **To cite this version:**

Nelida Mirabet-Herranz, Jean-Luc Dugelay. Beyond the visible: thermal data for facial soft biometric estimation. EURASIP Journal on Image and Video Processing, 2024, 2024 (1), pp.27. 10.1186/s13640-024-00640-5 . hal-04710175

**HAL Id: hal-04710175**

**<https://hal.science/hal-04710175v1>**

Submitted on 26 Sep 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Beyond the Visible: Thermal Data for Facial Soft Biometric Estimation

Nelida Mirabet-Herranz<sup>1\*</sup> and Jean-Luc Dugelay<sup>1</sup>

<sup>1</sup>Department of Digital Security, EURECOM, Campus SophiaTech, 450  
Route des Chappes, Biot, 06410, France.

\*Corresponding author(s). E-mail(s): [mirabet@eurecom.fr](mailto:mirabet@eurecom.fr);  
Contributing authors: [dugelay@eurecom.fr](mailto:dugelay@eurecom.fr);

## Abstract

In recent years, the estimation of biometric parameters from facial visuals, including images and videos, has emerged as a prominent area of research. However, the robustness of deep learning-based models is challenged, particularly in the presence of changing illumination conditions. To overcome these limitations and unlock new opportunities, thermal imagery has arisen as a viable alternative. Nevertheless, the limited availability of databases containing thermal data and the small amount of annotations on them limits the exploration of this spectrum. Motivated by this gap, this paper introduces the Label-EURECOM Visible and Thermal (LVT) Face Database for face biometrics. This pioneering database includes paired visible and thermal images and videos from 52 subjects along with metadata of 22 soft biometrics and health parameters. Due to the reduced number of existing databases in this domain, the LVT Face Database aims to facilitate further research and advancements in the utilization of thermal imagery for diverse eHealth applications and soft biometric estimation. Moreover, we present the first comparative study between visible and thermal spectra as input images for soft biometric estimation, namely gender age and weight, from face images on our collected database.

**Keywords:** Face database, Visible spectrum, Thermal spectrum, Gender, Age, Weight.

## 1 Introduction

Facial processing from visual content has gained significant attention in recent years, as the estimation of soft biometrics from faces has been proven important to support

various biometric systems [1, 2]. Moreover, remote estimation of health parameters from facial multimedia offers a non-invasive contactless estimation for assessing a subject’s health status, with applications ranging from medical emergencies and road accidents to at-home daily monitoring and telehealth.

Automatic face recognition has consistently been one of the most active research areas of computer vision [3]. Beyond people identification, the estimation of soft biometric traits such as gender and age, has been established in the literature. Moreover, a vast amount of soft biometrics and health information belonging to a subject has been proved to be embedded in face visuals [4]. The estimation of health indicators such as height, weight and Body Mass Index (BMI) from a single facial shot, has been explored in the literature by training a regression method based on the 50-layer ResNet-architecture [5]. Beyond physical attributes, researchers have extracted the called micro-signals from faces, information that has played important roles in media security and forensics [6]. The concept of remote photoplethysmography (rPPG) has evolved over the past fifteen years, utilizing the fact that blood absorbs more light than ambient tissues therefore subtle changes in blood volume can be captured by cameras based on the above-mentioned light absorption. This has allowed for remote photoplethysmography (rPPG). Research has shown how a mobile phone camera has enough resolution to capture rPPG signal from faces leading to a successful Heart Rate (HR) estimation [7]. Following the same principle, recent works have successfully approximated the Blood Pressure (BP) of a subject thanks to the difference between the times a pulse wave reaches two different parts of the face [8]. More recent investigations employ Convolutional Neural Networks (CNN) to compute the ratio of oxygenated to total hemoglobin (SpO2) from facial videos, considering direct and alternating current components extracted from RGB facial videos [9].

Facial processing models have traditionally based their estimations on images acquired in the visible spectrum. Despite these networks reaching a significant level of maturity with practical success, deep learning approaches relying on data from the visible spectrum are affected by compromising factors such as occlusion and changes in illumination. Thermal imagery has proven itself as a powerful capturing tool [3]. Computer vision researchers have affirmed its superiority to visible imaging in challenging conditions, such as the presence of smoke, dust, and the absence of light sources [10]. Thermal imagery operates by detecting electromagnetic radiation in the medium MWIR ( $3 - 8\mu m$ ) and long LWIR ( $8 - 15\mu m$ ) wave infrared spectrum [11], where skin heat is located. This capability enables thermal images to overcome the lack of illumination or certain types of occlusions. However, studies have highlighted how the thermal heat captured by thermal cameras can be influenced by factors such as ambient temperature or intense physical activity [3].

To move towards more accurate facial processing models and because we believe in the potential of thermal imagery, in this paper we are presenting a comparative study between visible and thermal spectra for the estimation of different soft biometrics in our new database. The main contributions of this work are as follows:

- We present our Label-EURECOM Visible and Thermal Face Database for face biometrics composed of 612 images and 416 videos from 52 different subjects and a compendium of 22 health metrics and soft biometrics annotated per person.

- We compare the performances of state-of-the-art deep learning-based models for the estimation of three soft biometric traits: gender, age, and weight when trained with visible versus thermal data.
- We offer insights into the strengths and weaknesses of each modality by testing the models with three types of conditions in our database: Studio lights, Presence of eyeglasses, and Ambient illumination.

The rest of the paper is organized as follows, Section 2 reviews existing works on facial processing for soft biometric and health parameters estimation and lists existing databases containing thermal visuals and some descriptors of them. In Section 3, our LVT Face Database for face biometrics newly collected is presented in detail. Section 4 includes a description of the networks studied as well as the evaluation protocol and metrics used in our experiments while in Section 5 the results of testing the three networks on our LVT database are discussed. Finally, Section 6 summarizes and concludes with the future directions of our work.

The LVT Face Database for face biometrics is publicly available upon request<sup>1</sup>.

## 2 Potential of visible and thermal paired data

Deep learning-based biometric systems and facial eHealth models are traditionally trained on databases acquired in the visible or, more recently, the Near InfraRed (NIR) spectra. In this Section, we present existing thermal databases as well as various studies that have focused on the thermal spectrum for facial processing applications such as cross-spectrum face recognition algorithms or HR estimation.

### 2.1 Relevant thermal databases

Interest in employing thermal face images has grown in the past years, nevertheless, its use has been mostly confined to tasks such as landmarks and face detection and FR [3, 12]. A relevant subset of FR is Cross-FR (CFR) discipline that aims to identify a person’s image in the thermal spectrum from a gallery containing face images acquired in the visible spectrum [13]. Only a few databases have been provided involving visuals acquired in thermal spectra, and among them, those covering soft and hidden biometric metadata are few. In Table 1, we present an exhaustive selection of relevant databases that include visuals in the thermal spectrum and some key descriptors of them including their year of release, the number of subjects, images and videos present in the database and their initial intended purpose.

One of the first datasets containing thermal visual data was presented in 2003 [14]. The data was acquired at the University of Notre Dame and contains images from 240 distinct subjects with four views showing different lighting and facial expressions with the purpose of recognizing individuals. Beyond people recognition, Wang *et al.* established a similar database for expression recognition, containing both spontaneous and intended expressions of more than 100 subjects [15], while Gault *et al.* recorded thermal videos from 32 subjects under three imaging scenarios and their paired rPPG signals for HR estimation [16]. In 2018, two new databases were acquired for FR with

---

<sup>1</sup><https://lvt.eurecom.fr/>

**Table 1** Relevant face databases containing visuals in thermal spectra.

Year	Dataset	# of subjects	# of images	# of videos	Objective
2003	UND-XI [14]	241	4584	-	FR
2010	NVIE [15]	215	Not provided	Not provided	Expression recognition
2013	TH-HR [16]	32	-	96	HR
2018	VIS-TH [3]	50	2100	-	FR
2018	TUFTS [17]	113	Over 10000	113	FR
2018	TH-HR-RR [18]	20	-	40	HR, RR
2021	Speaking faces [19]	142	-	45 hours	Biometric Authentication
2021	ARL-VTF [20]	395	549712	-	Cross-FR
2022	SF-TL54 [12]	142	2556	-	Landmarks detection
2023	Ours: LVT	52	612	416	Facial processing

multiple illumination, pose, and occlusion variations [3], including imagery from different modalities, namely visible, thermal, near-infrared, a computerized facial sketch, and 3D images of each volunteer’s face [17]. In the same year, Barbosa *et al.* collected thermal videos from 20 healthy subjects in two phases: phase A (frontal view acquisitions) and phase B (side view acquisitions), and the corresponding PPG and thoracic effort were simultaneously recorded for HR and Respiratory Rate (RR) estimation [18]. More recently, two large-scale visible and thermal datasets have been assembled. Abdrakhmanova *et al.* gathered a combination of thermal, visual, and audio data streams to support machine learning-based biometric applications [19], and Poster *et al.* presented the largest collection of paired visible and thermal face images to date. Variability in expression and pose were recorded [20]. Following, a thermal face dataset with annotated face bounding boxes and facial landmarks composed of 2556 images was introduced [12].

## 2.2 Thermal data in facial processing tasks

**Thermal data for Soft Biometrics:** Recent works have initiated the exploration of the thermal spectrum for facial processing models. In the literature, two soft biometrics, namely gender and ethnicity, have been estimated from thermal input data. In [21], the authors presented the first work on gender classification from faces in the thermal spectrum. They proposed a pipeline of techniques that consists of Local Binary Pattern method to detect the edges in an image, followed by Principal Component Analysis for dimensionality reduction, and finally, a Support Vector Machine as a classification technique for estimating the subjects’ gender. Similarly, Abouelenien *et al.* utilized the Eigenfaces method for visible faces and statistical measurements of pixel color for thermal faces, employing decision trees for gender classification. Fusion between visible and thermal data was integrated within the decision tree model [22]. More generally, works have explored gender estimation from nine different narrow spectral bands [23]. Deep learning structures began to be explored in [24], where the authors trained a VGG-CNN structure with visible faces and tested it on thermal faces for gender and ethnicity classification. Farooq *et al.* performed transfer learning from nine famous architectures to estimate gender from thermal data [25], including ResNet-50, ResNet-101, Inception-V3, MobileNet-V2, VGG-19, AlexNet, DenseNet-121, DenseNet-20, and EfficientNet-B44. They also proposed GENNet for the same

task. More recently, Abdrakhmanova *et al.* proposed a combination of bidirectional recurrent neural network and CNN for extracting features from Visible-Thermal-Audio and then performed fusion at the feature level to estimate the person’s gender [19]. They classify females and males afterward with a final decision layer.

**Thermal data for eHealth:** Although facial thermal imagery has traditionally been employed for face recognition tasks or gender estimation, some researchers have aimed to its use for eHealth parameter estimation in the thermal spectrum, highlighting the potential of this type of data. In 2017, Rai et al. suggested that thermal imaging systems have the ability to provide details about physiological processes through skin temperature distributions, influenced by factors such as blood perfusion. Thermal cameras are commonly employed in the medical field to observe minute temperature variations, with applications including the detection of malignant tumors [11]. The assessment of eHealth parameters such as heart rate from face videos has been studied in depth in recent years. To the best of the authors’ knowledge, all existing methods require proper illumination, posing challenges in uncontrolled environments. In 2018, Barbosa et al. introduced a novel method for remote HR monitoring based on periodic head movements resulting from the cyclical ejection of blood flow from the heart to the head. This innovative algorithm utilized thermal images as input data [18]. Furthermore, they demonstrated the feasibility of evaluating and measuring a subject’s respiratory rate by analyzing temperature fluctuations under the nose during the respiratory cycle. Thermal imagery has proven to be of high value in overcoming illumination constraints, given its light-invariant nature. In a similar line of research, ongoing works explore the potential of deep-learning approaches for extracting heart rate and blood pressure information from thermal images [26].

To the best of our knowledge, current literature has predominantly focused on extracting gender, ethnicity and eHealth parameters namely HR, RR and BP from thermal face data. The estimation of other traits such as age and weight from thermal images remains unexplored by the community. The motivation behind collecting a new database of visible face visual data and their thermal counterparts arises from the potential of using thermal images and videos as input data in facial processing tasks. Furthermore, existing databases are often limited to visual face information content and one or two biometric or health parameters. We believe that the value added by a database comprising more than 20 different soft biometric and health measures is a valuable contribution to the biometric community.

### 3 LVT Face Database description

In this section, we first introduce the recording setup of the database and the characteristics of the acquisition devices. We then elaborate on the data collection methodology, as well as the database design and the associated subjects’ metadata.

#### 3.1 Acquisition material

The visible and thermal face visual data were obtained using the dual sensor from the FLIR Duo R camera developed by FLIR Systems. This camera is specifically designed to capture visible and thermal visuals simultaneously, particularly suitable



**Fig. 1** Flir Duo R camera (left) and acquisition setup (right).

for unmanned aerial vehicles. The FLIR Duo R dual camera has been employed in recent research due to its appropriateness in data collection for various tasks such as face recognition and cross-spectrum applications [3, 10]. The visible and thermal sensors of this camera consist of a CCD sensor with a pixel resolution of  $1920 \times 1080$  and an uncooled VOx microbolometer with a pixel resolution of  $640 \times 512$ , respectively.

To assess the health status of the subjects, various devices were utilized. A contactless infrared thermometer with a precision of  $\pm 0.2^\circ\text{C}$  between  $34^\circ\text{C}$  and  $42.0^\circ\text{C}$  and a precision of  $\pm 0.3^\circ\text{C}$  in the range of  $42.1^\circ\text{C}$  and  $43.0^\circ\text{C}$  was employed for computing the user’s body temperature. For calculating BP, an OMRON HEM-7155-E tensiometer was used, along with a LED finger oximeter for SpO2 measurement with a precision of  $\pm 2\%$ . To track HR, subjects wore a Garmin Vivoactive®4 smartwatch equipped with an optical PPG sensor capable of detecting the heart rate by shining a green light through the subject’s skin, reflecting the red cells in the skin’s blood vessels. For quantifying bodyweight-related measures, the RENPHO®Body Fat Smart scale was utilized. When a subject steps on the device and enters their gender, age, and height into the system, the scale returns 13 metrics, including weight and BMI.

### 3.2 Visuals collection protocol

The image and video acquisition took place in an indoor environment with the ambient temperature set to  $25^\circ\text{C}$ . In Figure 1, we present the arrangement. The acquisition setup included a white wall serving as a background, and a chair positioned at a fixed distance of 0.25 m from the camera, which was placed at a height of 1 meter from the ground. Additionally, a two-point lighting kit was strategically placed to minimize shadows, facilitating the segmentation of the subject from the background. Each volunteer participated in two separate acquisition sessions, with an average time interval of 6 weeks. Before the acquisition process, volunteers were requested to fill out and sign consent forms.

The visual data comprises 6 images per person (3 visible and their associated thermal pair) in each session, encompassing three different conditions: Neutral (N),



**Fig. 2** Example images from the LVT Face Database. The three variations are displayed in visible (upper row) and thermal (bottom row) spectra, from left to right: N, O and A.

Ambient light (A), and an occlusion in the form of eyeglasses (O), resulting in a total of 612 images. Figure 2 illustrates example images of an individual from the database. Additionally, four 60-second videos were recorded per subject in each session under Neutral (N) conditions. The first pair of videos (one in the visible spectrum and its paired thermal counterpart) was captured after the subject had been resting for at least 5 minutes, while the second pair followed moderate exercise in the form of climbing stairs to elevate their HR values, resulting in a total of 408 60-second videos.

### 3.3 Subjects' metadata

Several pieces of metadata were collected to describe the subjects, including gender, age, and height. Additional parameters were quantified to assess their health status, such as body temperature, HR, BP, SpO<sub>2</sub>, weight, and BMI. In addition to weight and BMI, the smart scale provided 11 other variables: body fat and body water percentages, skeletal muscle, fat-free weight, muscle mass, bone mass, protein, subcutaneous and visceral fat, Basal Metabolic Rate (BMR), and metabolic age. The filenames for images and videos are constructed by indicating the visual data spectrum, subject ID, session ID (1 or 2), and in the case of images, the conditions at the time of acquisition (N, O, or A).

### 3.4 Summary

The presented database is designed as a collection of images, videos, soft biometrics, and health parameters recorded from 52 different subjects across two sessions. It consists of 612 face and shoulders images and 416 60-second videos, totaling approximately 285 GB of disk space. The 52 recorded participants, comprising 38 males and 14 females, are from 13 different countries spanning four continents, with ages ranging between 22 and 51 years. An executive summary of the dataset is provided in Table 2.



**Table 2** (Table is read horizontally) Summary of the information contained in the LVT Face Database.

<b>Identities</b>	52 subjects			2 sessions		
<b>Visuals</b>	6 paired images in conditions N, O A		2 paired 60s videos resting		2 paired 60s videos after activity	
<b>Metadata</b>	<b>Biometrics</b>	ID	Gender	Age	Height	Weight
	<b>Health paramet.</b>	Temperature	HR resting	HR activity	BP max	BP min
		BMI	Body fat(%)	Body water(%)	Muscle	Fat-free kg
		Body mass	Bone mass	Proteins	Subcutaneous fat	
		BMR	SpO2	Visceral fat	Metabolic age	

## 4 Methodology

In this section, we describe the models implemented and compared in our experiments, the evaluation metrics, and the experimental setup of the networks.

### 4.1 Soft biometric estimation models

VGGNet [27] was developed by the Visual Geometry Group from the University of Oxford to improve computer vision tasks by increasing the depth of an architecture with small convolutional filters of size  $3 \times 3$ . In addition to this, VGG incorporates  $1 \times 1$  convolutional layers to make the decision function more non-linear without changing the receptive fields. VGG architecture has been proven in the literature as powerful for estimating gender and age from face images [28]. Moreover, in their comparative study of architectures for gender estimation from thermal data, Farooq *et al.* [25] revealed the high performance of VGGNet for this task. No study, to the authors' knowledge, has been conducted on the feasibility of thermal imagery for age estimation. Therefore, we select the VGGNet network with 16 weight layers, i.e., the VGG16 model, for our gender and age estimation models. We use the VGG16 base model as a feature extractor and we add custom fully connected layers on top for binary classification and regression for gender and age prediction respectively.

Residual Neural Networks (ResNet) [29] are convolutional neural networks that introduce the concept of residual learning. Instead of learning a direct mapping between layers, ResNet learns the differences between the input and the desired output of a layer by using shortcut connections, also known as skip connections that bypass one or more layers and directly connect the input of a layer to its output. In the literature, face image-based weight estimation has been demonstrated using ResNet architectures with 50 layers and a final regression layer [5, 30, 31]. For our experiments, similar to those studies, we select a ResNet50 model.

### 4.2 Evaluation metrics

Accuracy is used as a metric for gender classifier assessment. Regarding age and weight, we report the Mean Absolute Error (MAE) and Mean Root Square Error (MSRE) in years and kilograms (kg) respectively and the Pearson's correlation coefficient ( $\rho$ ). Additionally, for age, we provide the Standard Deviation (StD) of the



**Fig. 3** Transfer learning protocol for soft biometric estimation from visible and thermal images.

difference between the predicted and the real age of the subjects. Finally, we include the Percentage of Acceptable Predictions (PAP) for the weight estimation network. This metric represents the percentage of predictions with an error smaller than 10% of the initial weight, indicating a reasonable error in medical applications.

### 4.3 Experimental setup

The VGG16 and ResNet50 architectures were implemented using TensorFlow and Keras frameworks. The weights of the VGG16 model were initialized with pre-trained weights obtained from the ImageNet dataset with its final fully connected layers excluded to add layers designed for the specific tasks under consideration. The output of the VGG16 base model was flattened and passed through a fully connected layer consisting of 256 neurons with ReLU activation. Additionally, a dropout of 0.5 was applied to prevent overfitting. For the binary gender classification task, a final output layer with a single neuron and sigmoid activation function was added, while for the age regression task, an output layer with a single neuron and linear activation function was incorporated. For the ResNet50 model, the weights were initialized with pre-trained weights obtained from the UTK dataset.

**Table 3** Accuracy of VGG-16 for gender estimation in thermal and visible spectra face images from the LVT test set.

<i>GENDER</i>	VIS			TH		
	N	O	A	N	O	A
Accuracy	<b>0.85</b>	<b>0.80</b>	0.75	0.80	0.75	<b>0.80</b>

**Table 4** Performance of VGG-16 for age estimation in thermal and visible spectra face images from the LVT test set.

<i>AGE</i>	VIS			TH		
	N	O	A	N	O	A
Std	7.08	7.24	6.82	<b>6.47</b>	<b>6.48</b>	<b>6.42</b>
MAE	5.77	6.20	5.53	<b>3.95</b>	<b>4.08</b>	<b>3.80</b>
MSRE	8.73	9.18	8.57	<b>7.31</b>	<b>7.44</b>	<b>7.25</b>
$\rho$	0.29	<b>0.32</b>	0.24	<b>0.35</b>	0.31	<b>0.39</b>

The input images were resized to  $224 \times 224$  pixels. Each network underwent two training sessions with identical configurations: one with visible images and another with thermal images, utilizing Transfer Learning (TL). A subject-exclusive split of the database was conducted, allocating 240 images for training and 60 for testing per spectra. The training and testing pipeline is illustrated in Figure 3.

The VGG16 networks were trained for 20 epochs using the Adam optimizer with a learning rate of 0.001. Binary cross-entropy loss function was selected for gender classification, while mean squared error was employed for age estimation. Each ResNet50 model was re-trained during 10 epochs followed by an additional 10 epochs for training the final regression layer. During each TL step, the first 20 layers were frozen. Adam optimizer was used with a learning rate of 0.01, and Huber loss function was selected with  $\delta = 1$ .

## 5 Experiments

In this section, we present a comparative study of the state-of-the-art networks for soft biometrics estimation using visible versus thermal images as input data. Additionally, we conduct a comparison of thermal and visible domains against various facial variations introduced in our database, reflecting the performance of both modalities in practical scenarios.

It has been proved that bone, muscle, and body fat do not conduct equally temperature [32]. Heat emission patterns can be used to characterize a person as they provide information about the location of major blood vessels, skeleton thickness, amount of tissues, and muscle and fat distribution<sup>2</sup>. Additionally, it is known that male and female bodies have different bone mineral and muscle density, and their facial appearance differs even when they are of the same weight [33]. Therefore we believe thermal imagery will access crucial information for the various soft biometric tasks considered in this research.

<sup>2</sup><https://biometrics.manguet.org/types/face.htm#thermogram>

**Table 5** Performance of ResNet50 for weight estimation in thermal and visible spectra face images from the LVT test set.

<i>WEIGHT</i>	VIS			TH		
	N	O	A	N	O	A
MAE	8.60	<b>9.01</b>	8.64	<b>6.84</b>	12.35	<b>7.05</b>
MSRE	11.01	<b>11.48</b>	10.86	<b>9.46</b>	14.91	<b>9.80</b>
$\rho$	0.33	0.38	0.38	<b>0.41</b>	<b>0.51</b>	<b>0.42</b>
PAP	40	<b>40</b>	40	<b>55</b>	20	<b>60</b>

In Table 3, we present the results of the VGG16 network trained on the visible (VIS) and thermal (TH) images of our dataset for gender classification. We observe that under natural conditions (N), where studio lights are used, visible data outperforms thermal imagery. This trend continues when occlusions in the form of eyeglasses (O) are present. In the thermal spectrum, glasses act as opaque barriers, leading to a loss of information in the eyes and penalizing this modality. However, in scenarios where no dedicated light sources are applied such as ambient light (A), VGG16 trained with thermal data delivers better results.

Table 4 displays the results of the VGG16 networks trained for age estimation. In contrast to gender estimation, thermal imagery exhibits clear superiority across all data variations. Even in the presence of eyeglasses, the error is only slightly higher compared to other variations.

Finally, Table 5 presents the results of the weight estimation network. Like age estimation, the metrics demonstrate that ResNet50 performs better in weight estimation when using thermal data, particularly in N and A conditions. This confirms the potential of thermal imagery in capturing hidden and detailed information from human faces, especially for age and weight estimation tasks using VGG and ResNet50 architectures, respectively.

## 6 Conclusion

This article introduces the LVT Face Database for face biometrics, containing visuals from 52 subjects captured under various conditions. The database comprises 306 visible and 306 thermal images, along with 204 visible and 204 thermal videos collected simultaneously using a paired camera (FLIR Duo R). This setup allows for the comparison or fusion of different data types. The visuals acquired are associated with metadata, covering both biometric and health-related information. To the best of our knowledge, this is the first database providing visible-thermal face images and recordings, accompanied by gender, age, body temperature, SpO2, blood pressure, heart rate (resting and after physical activity), height, weight, BMI, and 11 additional health metrics. The extensive annotations for each subject aim to unlock the potential of thermal data in assessing a person’s health status. Furthermore, experiments are conducted in this novel database demonstrating the feasibility of estimating three biometric traits: gender, age, and weight from facial thermal data. We partition the test set into three subsets based on the three variabilities presented in the LVT dataset: studio lights, occlusion in the form of eyeglasses, and ambient light. The results highlight the advantages of thermal imagery, especially for age and weight estimation

from faces, and demonstrate that thermal imaging is superior when no dedicated light sources are used, such as in ambient light conditions. Building on these promising results, future work will explore thermal imagery not only as an alternative but also as a complement to visible data. The estimation of other parameters, such as SpO<sub>2</sub> or height, from thermal depictions will be investigated.

## Declarations

### Availability of data and materials:

To allow other scholars to reproduce our results, we have made publicly available the LVT Face Database. More information can be found on the webpage <https://lvt.eurecom.fr/>. A download link for the database compressed and a password for decrypting the compressed LVT ZIP files will be provided after receiving the duly signed license agreement. Please fill in the license agreement and send a scanned copy by e-mail at [lvt@eurecom.fr](mailto:lvt@eurecom.fr).

### Competing Interests:

The co-author of this paper Jean-Luc Dugelay is the founder EiC.

### Author Contributions:

Nelida Mirabet-Herranz worked on the database collection, model implementation and results. Jean-Luc Dugelay provided assistance on the database and methodology design and supervision. Both authors worked on the final manuscript.

### Funding:

This work has been partially supported by the European CHIST-ERA program via the French National Research Agency (ANR) within the XAIface project (grant agreement CHIST-ERA-19-XAI-011).

### Acknowledgments:

Not applicable.

### Authors' information:

**Nelida Mirabet-Herranz** received her bachelor's degree in Mathematics from Universidad de Valencia in 2018 and her Master's degree in Data Science from EURECOM in 2020. She is currently a PhD student in facial biometrics and Deep Learning at EURECOM, Sophia Antipolis, under the supervision of Prof. Jean-Luc Dugelay and Prof. Chiara Galdi.

**Jean-Luc Dugelay** obtained his PhD in Information Technology from the University of Rennes in 1992. He is currently a Full Professor at EURECOM, Sophia Antipolis. His current work focuses on Image Engineering & Security. He has co-authored 300+ publications. He is the founding Editor-in-Chief of the EURASIP Journal on image and Video Processing (SPRINGEROPEN).

## References

- [1] Dantcheva, A., Velardo, C., D'angelo, A., Dugelay, J.-L.: Bag of soft biometrics for person identification. *Multimedia Tools and Applications* **51** (2011)
- [2] Jain, A.K., Dass, S.C., Nandakumar, K.: Soft biometric traits for personal recognition systems. In: *International Conference on Biometric Authentication*, pp.

731–738 (2004). Springer

- [3] Mallat, K., Dugelay, J.-L.: A benchmark database of visible and thermal paired face images across multiple variations. In: 2018 International Conference of the Biometrics Special Interest Group (BIOSIG), pp. 1–5 (2018). IEEE
- [4] Ross, A., Banerjee, S., Chowdhury, A.: Deducing health cues from biometric data. *Computer Vision and Image Understanding* **221**, 103438 (2022)
- [5] Dantcheva, A., Bremond, F., Bilinski, P.: Show me your face and i will tell you your height, weight and body mass index. In: 2018 24th International Conference on Pattern Recognition (ICPR) (2018). IEEE
- [6] Wu, M.: Exploiting micro-signals for physiological forensics. In: Proceedings of the 2020 ACM Workshop on Information Hiding and Multimedia Security, pp. 1–1 (2020)
- [7] Rahman, H., Ahmed, M.U., Begum, S., Funk, P.: Real time heart rate monitoring from facial rgb color video using webcam. In: The 29th Annual Workshop of the Swedish Artificial Intelligence Society (SAIS), 2–3 June 2016, Malmö, Sweden (2016). Linköping University Electronic Press
- [8] Lu, Y., Wang, C., Meng, M.Q.-H.: Video-based contactless blood pressure estimation: A review. In: 2020 IEEE International Conference on Real-time Computing and Robotics (RCAR), pp. 62–67 (2020). IEEE
- [9] Akamatsu, Y., Onishi, Y., Imaoka, H.: Blood oxygen saturation estimation from facial video via dc and ac components of spatio-temporal map. In: ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 1–5 (2023). IEEE
- [10] Eddine, M.J., Dugelay, J.-L.: Gait3: An event-based, visible and thermal database for gait recognition. In: 2022 International Conference of the Biometrics Special Interest Group (BIOSIG), pp. 1–5 (2022). IEEE
- [11] Rai, M., Maity, T., Yadav, R.: Thermal imaging system and its real time applications: a survey. *Journal of Engineering Technology* **6**(2), 290–303 (2017)
- [12] Kuzdeuov, A., Koishigarina, D., Aubakirova, D., Abushakimova, S., Varol, H.A.: Sf-tl54: A thermal facial landmark dataset with visual pairs. In: 2022 IEEE/SICE International Symposium on System Integration (SII), pp. 748–753 (2022). IEEE
- [13] Anghelone, D., Chen, C., Faure, P., Ross, A., Dantcheva, A.: Explainable thermal to visible face recognition using latent-guided generative adversarial network. In: 2021 16th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2021), pp. 1–8 (2021). IEEE

- [14] Kevin, X., Bowyer, W.: Visible-light and infrared face recognition. In: Workshop on Multimodal User Authentication, p. 48 (2003). Citeseer
- [15] 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 Transactions on Multimedia* **12**(7), 682–691 (2010)
- [16] Gault, T., Farag, A.: A fully automatic method to extract the heart rate from thermal video. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 336–341 (2013)
- [17] Panetta, K., Wan, Q., Agaian, S., Rajeev, S., Kamath, S., Rajendran, R., Rao, S.P., Kaszowska, A., Taylor, H.A., Samani, A., *et al.*: A comprehensive database for benchmarking imaging systems. *IEEE transactions on pattern analysis and machine intelligence* **42**(3), 509–520 (2018)
- [18] Barbosa Pereira, C., Czaplik, M., Blazek, V., Leonhardt, S., Teichmann, D.: Monitoring of cardiorespiratory signals using thermal imaging: a pilot study on healthy human subjects. *Sensors* **18**(5), 1541 (2018)
- [19] Abdrakhmanova, M., Kuzdeuov, A., Jarju, S., Khassanov, Y., Lewis, M., Varol, H.A.: Speakingfaces: A large-scale multimodal dataset of voice commands with visual and thermal video streams. *Sensors* **21**(10), 3465 (2021)
- [20] Poster, D., Thielke, M., Nguyen, R., Rajaraman, S., Di, X., Fondje, C.N., Patel, V.M., Short, N.J., Riggan, B.S., Nasrabadi, N.M., *et al.*: A large-scale, time-synchronized visible and thermal face dataset. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 1559–1568 (2021)
- [21] Chen, C., Ross, A.: Evaluation of gender classification methods on thermal and near-infrared face images. In: 2011 International Joint Conference on Biometrics (IJCB), pp. 1–8 (2011). IEEE
- [22] Abouelenien, M., Pérez-Rosas, V., Mihalcea, R., Burzo, M.: Multimodal gender detection. In: Proceedings of the 19th ACM International Conference on Multimodal Interaction, pp. 302–311 (2017)
- [23] Vetrekar, N., Naik, A., Gad, R.: Cross-spectral gender classification using multi-spectral face imaging. In: *Journal of Physics: Conference Series*, vol. 1921, p. 012048 (2021). IOP Publishing
- [24] Narang, N., Bourlai, T.: Gender and ethnicity classification using deep learning in heterogeneous face recognition. In: 2016 International Conference on Biometrics (ICB), pp. 1–8 (2016). IEEE
- [25] Farooq, M.A., Javidnia, H., Corcoran, P.: Performance estimation of the state-of-the-art convolution neural networks for thermal images-based gender classification

- system. *Journal of Electronic Imaging* **29**(6), 063004–063004 (2020)
- [26] Nair, K.S., Sarath, S.: Illumination invariant non-invasive heart rate and blood pressure estimation from facial thermal images using deep learning. In: 2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT), pp. 1–7 (2021). IEEE
- [27] Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. In: 3rd International Conference on Learning Representations (ICLR 2015) (2015). Computational and Biological Learning Society
- [28] Gyawali, D., Pokharel, P., Chauhan, A., Shakya, S.C.: Age range estimation using mtcnn and vgg-face model. In: 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), pp. 1–6 (2020). IEEE
- [29] He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
- [30] Mirabet-Herranz, N., Mallat, K., Dugelay, J.-L.: New insights on weight estimation from face images. In: 2023 IEEE 17th International Conference on Automatic Face and Gesture Recognition (FG), pp. 1–6 (2023). IEEE
- [31] Mirabet-Herranz, N., Dugelay, J.-L.: Lvt face database: A benchmark database for visible and hidden face biometrics. In: 2023 International Conference of the Biometrics Special Interest Group (BIOSIG), pp. 1–6 (2023). IEEE
- [32] Morley, M.: Thermal conductivities of muscles, fats and bones. *International Journal of Food Science & Technology* **1**(4), 303–311 (1966)
- [33] Han, D., Zhang, J., Shan, S.: Leveraging auxiliary tasks for height and weight estimation by multi task learning. In: 2020 IEEE International Joint Conference on Biometrics (IJCB), pp. 1–7 (2020). IEEE