2020 25th International Conference on Pattern Recognition (ICPR)
Milan, Italy, Jan 10-15, 2021

# Inner Eye Canthus Localization for Human Body Temperature Screening

Claudio Ferrari*, Lorenzo Berlincioni*, Marco Bertini, Alberto Del Bimbo
Media Integration and Communication Center
University of Florence, Italy
Email: {claudio.ferrari, lorenzo.berlincioni, marco.bertini, alberto.delbimbo}@unifi.it

***Abstract*—In this paper, we propose an automatic approach for localizing the inner eye canthus in thermal face images. We first coarsely detect 5 facial keypoints corresponding to the center of the eyes, the nosetip and the ears. Then we compute a sparse 2D-3D points correspondence using a 3D Morphable Face Model (3DMM). This correspondence is used to project the entire 3D face onto the image, and subsequently locate the inner eye canthus. Detecting this location allows to obtain the most precise body temperature measurement for a person using a thermal camera. We evaluated the approach on a thermal face dataset provided with manually annotated landmarks. However, such manual annotations are normally conceived to identify facial parts such as eyes, nose and mouth, and are not specifically tailored for localizing the eye canthus region. As additional contribution, we enrich the original dataset by using the annotated landmarks to deform and project the 3DMM onto the images. Then, by manually selecting a small region corresponding to the eye canthus, we enrich the dataset with additional annotations. By using the manual landmarks, we ensure the correctness of the 3DMM projection, which can be used as ground-truth for future evaluations. Moreover, we supply the dataset with the 3D head poses and per-point visibility masks for detecting self-occlusions. The data is publicly available at https://www.micc.unifi.it/resources/datasets/thermal-face/.**

## I. Introduction

When properly implemented, infrared thermography is the most efficient non-contact, cost effective and reasonably accurate solution for mass screening of individuals for elevated body temperature; this type of screenings is commonly used in public environments like airports, malls, hospitals and has become a way to non intrusively detect body fever, a common precursor of diseases like H1N1 flu, SARS, MERS and COVID-19 [1]–[3]. Such screening is normally conducted by measuring the skin temperature on a subject's face. However, not all the face regions are suitable for this task; it is widely recognized that the inner eye canthus represents the most stable region where to obtain a reasonably reliable temperature measure [4], [5]. This because, in normal conditions, it is both the warmest face region, and the most invariant to environmental factors that can alter the skin superficial temperature and, unlike the forehead, is not influenced by exertion and physical activity.

However, accurately detecting such regions on thermal imagery is a very challenging task as: (i) even slight head pose changes will result in self-occlusions and consequent obstruction of the canthus; (ii) the majority of thermal face databases do not contain annotations or, in case they do, they do not explicitly account for those regions, making it hard to both develop and evaluate a detection algorithm.

In this paper we propose a method for automatic detection of the inner eye canthus region based on coarse detection of 5 facial keypoints corresponding to eyes, nose and ears in thermal images, and on their 2D-3D mapping on a 3D morphable face model (3DMM) [6]. The 3DMM is a statistical shape modeling method that has been widely employed in many applications, from biometrics [7], [8], graphics [9], to reconstruction [10], [11] or even medical imaging [12]. This process has several benefits over direct region localization, since it allows to obtain additional information regarding the pose, i.e. understanding if the face is directly looking at the camera and estimating self-occlusions of face parts.

A second contribution of this work is the release of additional face annotations for a thermal face images dataset, including 3D head pose and face regions visibility masks.

The paper is organized as follows: related works are discussed in Sect. II, the proposed method is described in Sect. III; the dataset used and the additional annotations created to extend it are presented in Sect. IV, while experimental results are presented in Sect. V. Finally conclusions are drawn in Sect. VII.

## II. Related Work

The scientific literature on visible spectrum face images is extremely vast; even considering only works related to facial landmarks detection and pose recognition it is better to refer the reader to surveys like [13], [14]. In recent years, the improvement in resolutions and decrease in costs of thermal cameras has led to the development of methods specifically designed for the analysis of images in this spectrum.

*Thermal face datasets:* Due to the difficulty in creating them, compared to visible spectrum datasets, thermal face datasets are much smaller both in terms of number of identities (typically a few tens) and images (typically a few hundreds or thousands).

The PUCV Thermal Temporal Face (PUCV-TTF) dataset [15] consists of thermal images taken over time, so to study temporal changes in thermal imaging. It includes 46 people with five subsets for each subject, and each subset has 50 images.

*Both authors contributed equally to this research.

978-1-7281-8808-9/20/$31.00 ©2020 IEEE


2020 25th International Conference on Pattern Recognition (ICPR) | 978-1-7281-8808-9/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICPR48806.2021.9412015

The UND Collection X1 dataset [16] is composed of $\sim$ 2300 images of 82 subjects, and contains both high resolution visible images and corresponding low resolution ($320 \times 240$ pixels) thermal images.

The UL-FMTV dataset [17] of high resolution thermal videos has been obtained from 238 subjects over four years, considering pose, occlusions, time lapse.

In [18] a high-resolution thermal dataset of faces, filmed in different poses and expressions has been provided; it is composed of $\sim$ 3000 images from 90 subjects. Faces have been manually annotated with 68 keypoints.

*Thermal image analysis:* In general, thermal images have proven a challenging obstacle for many computer vision and image analysis tasks, like face recognition and re-identification [19], face part detection, keypoint localization, etc., due to their low contrast, low resolution, and lack of texture information.

In [20] has been presented one of the first CNN architectures to perform face recognition in the thermal domain only; in [15] local engineered features like SIFT, SURF, LBP have been used to evaluate the performance of face recognition under varying temporal conditions that affect thermal imaging, like environmental conditions and physiological changes that may happen over time. Face recognition can be hampered by the presence of glasses, that result in facial occlusion in thermal imagery; to deal with such occlusions the method presented in [21] uses a bag of CNN features model.

In [22] the authors evaluate how different deep neural network architectures for keypoint localization adapt to thermal face imaging, in order to perform thermal-to-visible face alignment and recognition. The results show that learning a global face appearance reduces critical localization errors.

In [23] the authors of the dataset originally proposed in [18] evaluate the performance of different keypoint localization with deep neural network comparing this approach with SIFT and HOG, then perform face expression recognition. The same authors have proposed in [24] a modular system for face detection, face tracking, head pose estimation, and emotion recognition using HOG and SIFT features.

In [25] is proposed to use a multi-task neural network to jointly consider facial landmark detection and emotion recognition for thermal face images. The network is composed by two parts: the first one is based on the U-Net structure, to extract good features, the second part of the network contains two branches, one for landmark detection and the other for emotion recognition.

*Thermography:* Face thermography is a useful new technique for psychophysiological research and medical applications; unlike other physiological measures it is uniquely contact-free, allowing its deployment in many real-world scenarios, like airports, hospitals, etc.

In [28] the authors have analyzed the influence of angles and distance on assessing temperature measurement using inner-canthi of the eye and thermal cameras, finding that when the face is not directly facing the camera, e.g. rotating more than 30° may result in a measure difference of 1-2°C.

In [29] the authors propose to use variation in temperature measurements of different regions of the face to recognize the presence of mental stress; the authors find that signals extracted from the upper lip region correspond well with high stress levels, while no correspondence can be shown for the other regions like forehead and nose. Use of thermographic measurement from faces of students has been proposed to assess learning difficulty in digital learning environments in [30]. In [2] the authors use both a task-constrained deep convolutional network pretrained to detect 5 keypoints and then to detect 68 keypoints, in combination with OpenPose, to recognize face landmarks in thermal images. These landmarks are then used to identify 4 ROIs (forehead, nose tip, right and left cheeks) used to measure skin temperature, in order to detect temperature changes associated with auditory stimulus.

Fever screening using termography has received a large attention following previous pandemic outbreaks like the severe acute respiratory syndrome (SARS) [31] and H1N1 flu (also known as "swine-flu") [5]. In these works the best correspondences between face parts and core body temperature have been analyzed resulting in guidelines and international standards on how to perform the screening, e.g. selecting the inner canthi area of the eyes [5], [32].

In order to automatize the process a few works have dealt with eye localization in thermal images. In [33] the Randomized Hough Transform has been used to localize the ellipses of the eyes and thus their inner canthi. The method has been tested on a dataset of 125 faces with a maximum angle deviation of 15° from the frontal pose. In [34] continuous body temperature measurements is obtained using a simple geometric model to detect forehead in thermal images and fine tuning a SSD detector with MobileNet backbone to perform face detection in thermal images. In [35] RGB and thermal imagery is combined to compute face landmarks in the visible spectrum and then obtain thermometry information with the thermal camera, to assess comfortable indoor conditions. Thermometric information obtained from a set of keypoints localized next to veins and capillaries has been proposed in [36] to perform face recognition.

## III. Proposed Method

The proposed algorithm consists of three main steps:

- Coarse detection of 5 keypoints *i.e.* eyes, nosetip and ears;
- 3D head pose estimation, model projection, and localization of the inner eye canthus region;
- Eye canthus location refinement.

The 5 initial keypoints are detected by means of an "off the shelf" detector *i.e.* OpenPose (OP) [26]. Despite being designed for visible spectrum imagery, it can robustly detect body, face, and hands keypoints rather accurately also on thermal images. In this work, we employ only the 5 facial keypoints, discarding all the others. As 3D face model, we employ the DL-3DMM proposed by Ferrari *et al.* [27], [37]. The DL-3DMM comprises a generic 3D face model $\mathbf{m} \in \mathbb{R}^{3\times6704}$ with 6704 vertices, and a set of deformation components $\mathbf{C} \in \mathbb{R}^{300\times6704}$ that allow deforming the generic model.

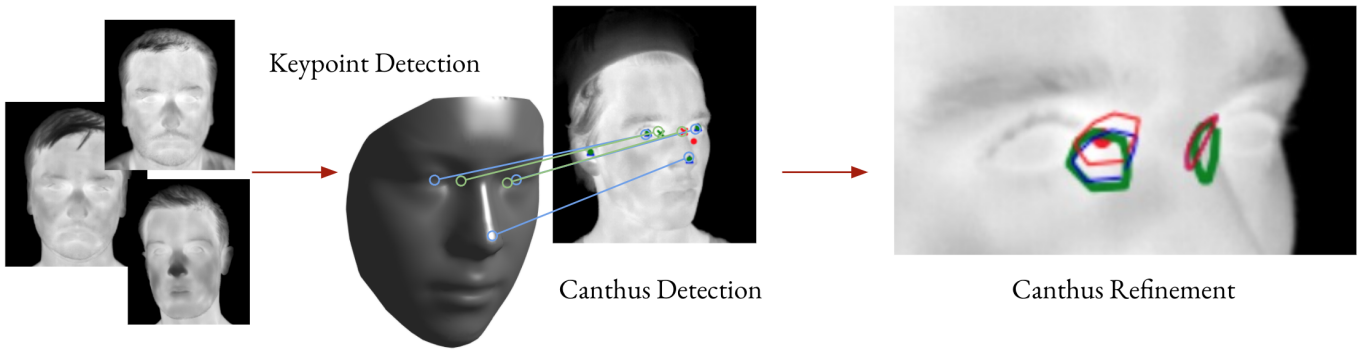



Fig. 1. Proposed method overview. Given a thermal face image, we first detect 5 keypoints (center of the eyes, ears, and nosetip) using the OpenPose detector [26]; then, we employ a 3DMM [27] to estimate the head pose and project the 3D face model on the image plane exploiting 2D-3D correspondence of points (blue lines). We subsequently locate the eye canthi by means of the 3D model (green lines). Finally, ( we refine the canthi localization (red point) by searching for the warmest area around the estimated canthi (green circles).

In the following, we separately describe the two subsequent steps *i.e.* pose estimation plus model projection, and the eye canthus localization refinement. In Fig. 1, the whole framework is illustrated.

### A. Head Pose Estimation and Model Projection

In order to estimate the 3D pose and a 3D to 2D projection exploiting the OP keypoints, we need to annotate a corresponding set of points on the 3D model. Let $\mathbf{l}_{op} \in \mathbb{R}^{2\times 5}$ and $\mathbf{L}_{op} \in \mathbb{R}^{3\times 5}$ be the 2D detected and 3D annotated keypoints, respectively. We estimate the projection using the orthographic camera model:

$$\mathbf{l}_{op} = \mathbf{A} \cdot \mathbf{L}_{op} + \mathbf{t} \ , \tag{1}$$

where $\mathbf{A} \in \mathbb{R}^{2\times 3}$ contains the affine camera parameters, and $\mathbf{t} \in \mathbb{R}^{2\times 5}$ is the translation on the image. To estimate the parameters, firstly, we recover the affine matrix $\mathbf{A}$ by solving the following least squares problem:

$$\arg\min_{\mathbf{A}} \|\mathbf{l}_{op} - \mathbf{A} \cdot \mathbf{L}_{op}\|_2^2 \ , \tag{2}$$

for which the solution is given by $\mathbf{A} = \mathbf{l}_{op} \cdot \mathbf{L}_{op}^{+}$, where $\mathbf{L}_{op}^{+}$ is the pseudo-inverse matrix of $\mathbf{L}_{op}$. We can employ a simple least squares solution since, by construction, OpenPose assumes a consistent structure of the 3D face parts, so not permitting unreasonable keypoint arrangements (e.g., the nose will never be detected above the eyes). Finally, the 2D translation can be estimated as $\mathbf{t} = \mathbf{l}_{op} - \mathbf{A} \cdot \mathbf{L}_{op}$. The estimated projection $\mathbf{P} = [\mathbf{A}, \mathbf{t}]$ is used to map each vertex of the model $\mathbf{m}$ onto the image.

As discussed previously, even slight head rotations can lead to the self-occlusion of the eye canthus regions, being them located in the nose side concavity. Thus, in order to be able to accurately perform a reliable temperature measurement at that points, we need to detect whether they are visible o not, totally or even partially. To this aim, we extract the 3D rotation matrix $\mathbf{R} \in \mathbb{R}^{3\times 3}$ from the affine projection matrix $\mathbf{A}$ by means of QR decomposition. This is possible thanks to the orthogonality property of rotation matrices *i.e.* $\mathbf{R}^T = \mathbf{R}^{-1}$. We then use the rotation $\mathbf{R}$ to rotate $\mathbf{m}$ according to the estimated pose, and calculate the visibility of each 3D vertex from the novel vewpoint using the method proposed by Katz *et al.* [38]. This way, we can easily detect which face parts are self-occluded.

The eye canthus region, as well as any other face point, can now be detected on the face image first by annotating the points of interest in the 3D model, and subsequently calculating their corresponding 2D coordinates by means of the projection matrix $\mathbf{A}$. Note that the annotations must be done just once. An example is shown in Fig. 2; the green crosses represent the projection of the eye canthus that were manually annotated on the 3D model and projected onto the image. Note also that when the subject is not frontal, we can detect the self occlusion. This aspect is important since OpenPose (as well as the majority of landmark detectors) do not usually provide indication of occlusions; rather, they either attempt to approximate the localization even if points are not directly visible, or do not return the detection at all.

### B. Eye Canthus Refinement

One limitation of the proposed approach is that it heavily relies on the accuracy of the keypoint detections. In order

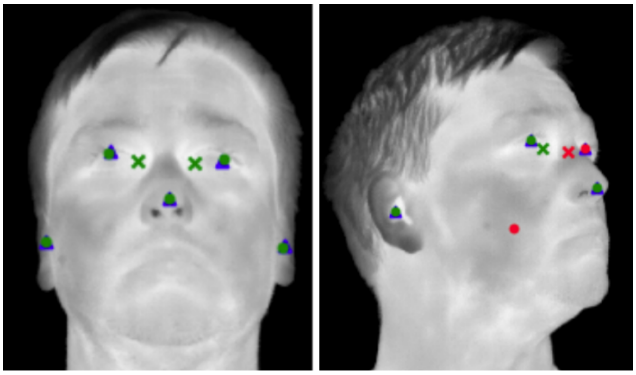


Fig. 2. Visualization of OpenPose keypoints (blue triangles) and our estimated projections (green/red dots). Red color indicates the projected keypoint is marked as occluded. The green crosses are our projected eye canthus points, which are not detected by OpenPose.

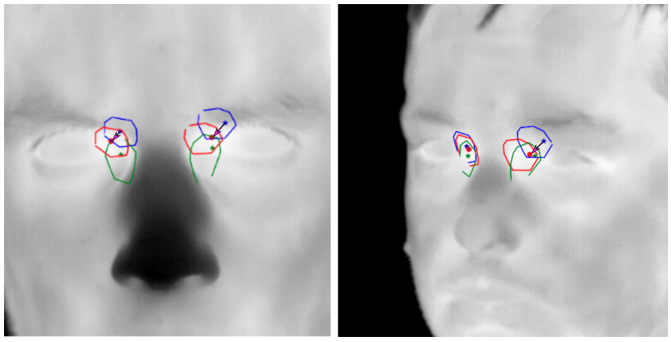


Fig. 3. Examples of eye canthus region refinement. In green manually annotated chantus and region, in blue OpenPose estimation and in red the refinement result.

to account for possible inaccuracies, we refine the canthus localization by searching for the warmest point in a neighborhood. This is motivated by the assumption that, locally, the eye canthus and its surrounding is expected to be the warmest face area [5]. So, we first annotate the two eye canthus $\mathbf{L}_{ch}$ in the 3D model $\mathbf{m}$, and define a region around them; each region is defined as the convex hull of the k-ring of $\mathbf{L}_{ch}$ (k-ring is the set of vertices that are distant from $\mathbf{L}_{ch}$ at most $k$). Within this area, we take the hottest point, *i.e.* the brightest pixel, and consider it as the new center of the eye canthus. An example of this process is shown in Fig. 3. Prior to computing the hottest point, we perform a Gaussian smoothing to account for sensor noise.

This strategy has a two-fold advantage: first, it allows correcting possible inaccurate detections; second, it ensures the measurement of the maximum temperature, representing at all effect a safe upper bound. In terms of thermal screening, this allows preventing dangerous false negatives.

## IV. Validation Data

In this section, we describe the dataset used for validating our approach, and how we enriched such data to allow quantitatively assessing the localization accuracy. We make use of the FaceDB dataset collected by Kopaczka *et al.* [18]. It contains 2935 high resolution ($1024 \times 768$ pixels) thermal frames of 90 subjects; for each subject, various sequences are recorded including pose variations, expression variations and basic action units activation. This dataset is one of the few including manually annotated landmarks; more precisely, each face image is annotated with 68 landmarks following the standard configuration as used in other datasets, such as the Helen dataset [39] (see Fig. 4). In this particular configuration, some of those landmarks do cover the contour of the eyes, including the eye canthus; however, we argue that, if not properly instructed to locate the canthus, different persons could interpret its location slightly differently. Moreover, as previously discussed, the ISO/TR 13154 termograph screening standard guidelines suggest to perform the measurement considering a small region rather than a single point, as reported in [5]. Thus, it is important to correctly identify such region, and ground-truth annotations are missing in this case.

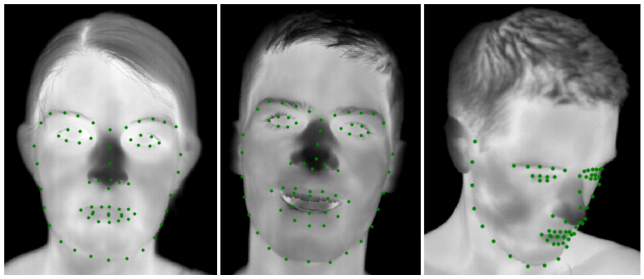


Fig. 4. Examples of the manually annotated landmarks in the dataset [18].

To address this, we propose to use our solution to augment the dataset and provide additional ground-truth annotations, that are derived automatically. In particular, we make use of the manually annotated landmarks to:

- Estimate the 3D pose and project the 3D model;
- Deform the 3D model to more accurately fit the face images, accounting for facial expressions;
- Identify the eye canthus regions by manually annotating a set of points in the 3D model and localizing the corresponding area on the images by means of the estimated projection.

The first step is performed as described in Section III-A, this time considering all the 68 available landmarks, which are assumed to be all correct and accurate. Next, we deform the 3DMM to fit the landmark locations, thus adapting its shape to each face image. While this is not possible with the OpenPose detections, that includes only 5 keypoints, the available 68 landmarks are sufficient to obtain a reasonable approximation of the face surface (see Fig. 4). To this aim, we exploit the method proposed in [40], which is fast and can account for strong facial expressions. Such method performs a geometric-based 3DMM fitting; it deforms the generic model $\mathbf{m}$ with the goal of minimizing the reprojection error between the annotated 2D landmarks and those obtained by projecting the 3D landmarks onto the image. To this aim, we estimate the 3DMM deformation coefficients $\boldsymbol{\alpha}$ solving the following:

$$\min_{\boldsymbol{\alpha}} \|\mathbf{l}_{gt} - \mathbf{P}(\mathbf{L}_{gt} - \mathbf{C}\boldsymbol{\alpha})\|_2^2 + \lambda \left\|\boldsymbol{\alpha} \circ \boldsymbol{\mu}^{-1}\right\|_2 . \tag{3}$$

where $\mathbf{l}_{gt}$ and $\mathbf{L}_{gt}$ are the 2D ground-truth landmarks and the 3D landmarks respectively, $\mathbf{P}$ indicates the projection onto the image, and $\boldsymbol{\mu}$ is a regularization term. To solve the problem we first pre-compute the landmarks displacement $\mathbf{X} = \mathbf{l}_{gt} - \mathbf{P}\mathbf{L}_{gt}$ and the deformation components projection $\mathbf{Y} = \mathbf{P}\mathbf{C}$. The latter is necessary to be able to fit the landmark locations directly on the image plane, The solution can be then found in closed form as:

$$\boldsymbol{\alpha} = \left(\mathbf{Y}^T\mathbf{Y} + \lambda \cdot \mathrm{diag}(\hat{\boldsymbol{\mu}}^{-1})\right)^{-1} \mathbf{Y}^T\mathbf{X} , \tag{4}$$

For more details on the solution of the minimization problem, the reader can refer to [27]. The average model $\mathbf{m}$ is then deformed into a new shape $\mathbf{S}$ as follows:

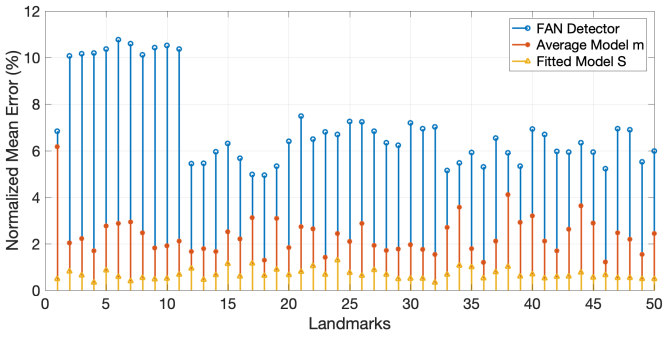



Fig. 5. Landmarks detection/reprojection error. In this configuration, the left/right eye canthi are the landmarks 22 and 25, respectively.

$$\mathbf{S} = \mathbf{m} + \sum_{i=1}^{k} \mathbf{C}_i \alpha_i \ . \tag{5}$$

This step is fundamental to obtain a consistent annotation across the images. In fact, if not properly modeled, facial expressions would eventually induce a misalignment between the reprojected points and compromise the pose estimation accuracy. The deformed 3D model $\mathbf{S}$ is then projected onto the image and used to annotate the eye canthi regions; in particular, similarly to Section III-B, the convex hull defined by the k-ring of the eye canthus ($k = 1, \ldots, 4$) is used as ground-truth.

Finally, using the same strategy as expounded in Section III-A, we estimate a dense visibility map for each deformed model $\mathbf{S}$, so that each image of the dataset is enriched with such information.

## V. Experimental Results

We validate the proposed eye canthus detection approach with an extensive set of experiments. In this regard, we previously discussed that, for this particular task, performing a quantitative accuracy assessment is made difficult by the partial lack of ground-truth data. So, we first validate the new ground-truth data that we generated as expounded in Section IV. Then, using this new data, we report detection accuracy results for the proposed eye canthus detection.

### A. Ground-truth validation

In this section, we provide a quantitative assessment of the additional ground-truth data. It is important that the new data which is intended to be used as ground-truth is correct. To this aim, we evaluate the landmarks reprojection error with respect to the manual annotations contained in the dataset, in order to ensure our annotations are meaningful. We also show that the 3DMM deformation step is important to account for the presence of facial expressions that can impair a correct pose and projection estimation. In Fig. 5, the Normalized Mean Error (NME) is reported separately for each landmark. The NME is a standard measure in the field of landmark detection and is computed as the Euclidean distance between the ground-truth and detected landmarks, normalized by the square root of the face bounding box size. Since no annotations are provided for such bounding box, we used an approximated box defined by the external face contour landmarks. Note that we do not use the latter landmarks for neither estimating the pose nor fitting the 3DMM; this because defining a unique face contour in 2D is not possible due to self-occlusions, which make the definition of the face silhouette inconsistent across 2D and 3D [41].

From Fig. 5, we can see that adapting the 3DMM is fundamental to obtain accurate model projections. We also reported the error obtained using the state-of-the-art FAN landmark detector [42] as baseline. Compared to that, we argue that our annotations can be considered sufficiently accurate to be used as ground-truth for further evaluation.

### B. Eye canthus region detection

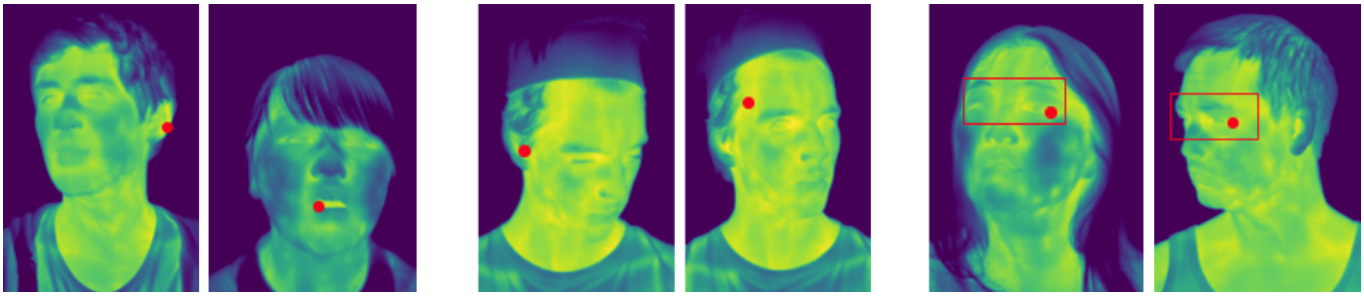
We report here a quantitative evaluation of the eye canthus detection accuracy, performing three different experiments. We first evaluate the detection accuracy with respect to the eye canthus landmarks in terms of NME; second, we use our annotations to compute the intersection over union (IoU) between the ground-truth and detected canthus regions. Finally, we also use the additional visibility masks to assess the accuracy of the occlusion detection. Results are reported in Table I. NME errors are computed both with respect to the manual annotations contained in the dataset (man) and our proposed annotations (gt); we can observe that the two measures are accurate and very similar, again confirming the validity of the proposed data. The IoU measure instead suggests that using a larger region can be convenient to better approximate the correct area. We can observe we also correctly detect whether the canthi are occluded or not in the 90% of the cases. The occlusion detection is formulated as binary decision (occluded / non-occluded ) applied to both the eyes, using the estimated visibility map.

In all the cases, we witness a slight drop of accuracy after the warmest point-based refinement. We already discussed that manual annotators could have not been properly instructed to label the eye canthi for the particular goal of performing thermal screening, and this could be a result of that. However, in this particular case, looking for the warmest point is beneficial in as much as it prevents under-estimating the temperature, which could potentially lead to dangerous false negatives.

### C. Head Pose Estimation

We evaluate the accuracy of the head pose estimation in terms of head rotation angles (Euler angles), that is pitch angle $\alpha$ ($x$-axis), yaw angle $\beta$ ($y$-axis), and roll angle $\gamma$ ($z$-axis or in-plane rotation). The three angles are extracted from the rotation matrices $\mathbf{R}$ obtained using the OpenPose keypoints, and compared with the ones provided with the additional ground-truth data. The error is computed as the absolute value of the angle differences (in degrees), that is

$$(e_\alpha, e_\beta, e_\gamma) = (|\alpha_{gt} - \alpha_{op}|, |\beta_{gt} - \beta_{op}|, |\gamma_{gt} - \gamma_{op}|)$$



(a) (b) (c)

Fig. 6. Examples of wrong detections when searching for the hottest point. Wrong detections can be caused for example by environmental factors (a), or changes in pose that influence the skin measurement (b). Even if restricting the search to the eyes region (c) wrong detection still occur.

TABLE I
EYE CANTHUS DETECTION ACCURACY (* MEANS RESULT IS THE SAME)

| | | FAN [42] | OP+3DMM | Refinement |
|---|---|---|---|---|
| 1-Ring | IoU | - | 16.5 ± 3.5 | 8.1 ± 2.1 |
| | NME (man)(%) | 6.8 * ± 2.1 | 4.1 ± 0.3 * | 5.6 ± 1.3 |
| | NME (gt)(%) | 6.5 * ± 2.2 | 3.7 ± 0.3 * | 4.9 ± 1.3 |
| 2-Ring | IoU | - | 32.5 ± 4.8 | 23.4 ± 3.6 |
| | NME (man)(%) | 6.8 *± 2.1 | 4.1 ± 0.3 * | 5.1 ± 1.1 |
| | NME (gt)(%) | 6.5 *± 2.2 | 3.7 ± 0.3 * | 4.5 ± 1.1 |
| 3-Ring | IoU | - | 41.7 ± 4.5 | 34.6 ± 2.8 |
| | NME (man)(%) | 6.8 *± 2.1 | 4.1 ± 0.3 * | 4.8 ± 1.2 |
| | NME (gt)(%) | 6.5 *± 2.2 | 3.7 ± 0.3 * | 4.3 ± 1.1 |
| 4-Ring | IoU | - | 47.1 ± 4.5 | 39.8 ± 2.8 |
| | NME (man)(%) | 6.8 *± 2.1 | 4.1 ± 0.3 * | 4.8 ± 1.1 |
| | NME (gt)(%) | 6.5 *± 2.2 | 3.7 ± 0.3 * | 4.4 ± 1.1 |
| | Occlusion (%) | - | 89.9 | - |

and averaged across all the images. The average error obtained is ($e_\alpha$, $e_\beta$, $e_\gamma$) = (8.98 , 7.16 , 6.62 ). To the best of our knowledge, the only previous work reporting head pose estimation results in the thermal domain is the one of Yu *et al.* [43], that estimates the head pose with an error $< 10°$ only in 17% of the cases. Kopaczka *et al.* [44] also developed a pose estimation module for thermal imagery, but evaluate their classifier on sequences in the visible spectrum. However, recent state-of-the-art algorithms that estimate the pose from single RGB images obtain a general error around 5° in challenging "in the wild" scenarios, *e.g.* [45]. Overall our method is rather accurate, and can be at the very least used in practical scenarios to detect whether a subject is facing the camera or not.

### D. Warmest face point ambiguity

Thermal screening standards suggest to perform the temperature measurement in correspondence of the eye canthus as it is the region that best approximates the internal body temperature, thus being the warmest face area. Hence, one could argue that detecting the face and then searching for the hottest point could suffice to recover their location. We instead empirically found that this strategy would lead to many wrong detections. Some qualitative examples of detection errors are shown in Fig. 6. This strategy often fails because, first, the skin surface is sensible to environmental conditions and can significantly change its temperature in a short time frame; additionally, head pose changes can also alter the measurement [28], as well as wearing accessories. In Fig. 6 (c) we also tried restricting the search space to the eyes region, using the manually annotated landmarks to ensure a correct localization of the eyes. Also in this case, such search can result in wrong detections, often when the subject is not correctly facing the camera. This result suggests that: (i) accurately localizing the eye canthus region is important to prevent detecting a wrong measurement point; (ii) it is likewise important to detect when the subject is not facing the camera.

TABLE II
EXECUTION TIME OF EACH STEP OF THE ALGORITHM (FPS / SEC)

| OpenPose | 3D Pose | Visiblity | Refinement | Tot |
|---|---|---|---|---|
| 22 / 0.04 | 1K / 0.001 | 22 / 0.04 | 33 / 0.03 | 9 / 0.11 |

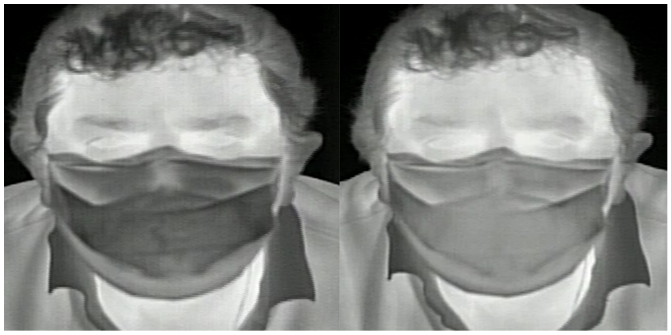
### E. Computational Time

Our approach is composed of three main steps, for which we report the execution time separately in Table II. The most computationally onerous steps are the OpenPose detection and the visibility mask computation, which both run at approximately 22 FPS (on a GTX 1080-Ti GPU, and Core i-7 CPU, respectively). The whole pipeline runs at approximately 9 FPS. Note that this is a prototype evaluation, and further optimizations can speed up the process.

In practical screening scenarios, time constraints are strict, and computing a per-point visibility mask might not be strictly necessary. In such cases, one can either choose to estimate the face pose and advise when the subject is not frontal *e.g.* fixing a threshold, or pre-compute an arbitrary number of masks corresponding to a discretized and fixed set of head rotations. This way the computational burden is noticeably reduced.

## VI. LIMITATIONS AND FUTURE WORK

The proposed approach can fairly accurately locate the eye canthus by refining the detections of an off-the-shelf algorithm. Still, there are some aspects that need further consideration. For example, when a subject wears eyeglasses, eyes are hidden by a black region; this happens because infrared sensors cannot



Inhaling Exhaling

Fig. 7. Example of temperature dynamics when wearing a surgical mask.

measure the temperature on reflective surfaces. If the detector ( OpenPose in this case ) is robust enough to return an estimate of the eyes position, then our approach still will provide an approximate location of the canthus. In this scenario, the temperature measurement would be invalid (the temperature would be out of a reasonable range) and an alarm can be raised. However, the best solution would be that of detecting the presence of glasses a priori with a glasses detector, and we will address this issue in future developments.

Another factor that should not be ignored is the use of surgical masks. During a pandemic, wearing a mask may be a strict requirement in crowded or indoor places so this aspect should not be disregarded. In fact, we observed that wearing a surgical mask, differently from other accessories like hats or scarfs, can alter the temperature that is measured at the eye canthus (see Fig. 7). This happens because cold/hot air flux generated when inhaling/exhaling or talking vents out directly around the eyes, and implies a strong temperature dynamics. This suggests collecting the measurement considering a short time frame, and that safety protocols should take into account this fact.

Finally, we couldn't perform real temperature measurements as no dataset is publicly available that contain per-point temperature information. In the future, we aim at collecting a dataset including such information for carrying a deepened analysis and validation of our approach.

## VII. Conclusion

In this paper, we have proposed an approach for detecting the eye canthus regions for human body temperature screening. Non-intrusive and non-contact thermography represents a viable and effective solution for mass screening of individuals for elevated body temperature, a common precursor of diseases. Our proposed approach works without employing images in the visible spectrum, thus preventing privacy related problems and allowing to use thermal-only cameras that have a reduced cost w.r.t. bi-spectrum cameras. Our approach can accurately detect the eye canthi also in presence of facial expression and head rotations. In addition, it can reliably detect when the canthus regions are self-occluded and when the subject is not facing the camera, which can impair the temperature measurement. Furthermore, we provided additional annotations that are publicly released to an existing dataset, so as to ease future evaluations.

## Acknowledgement

This research was partially funded by Leonardo.

## References

[1] A. Wilder-Smith, C. J. Chiew, and V. J. Lee, "Can we contain the COVID-19 outbreak with the same measures as for SARS?" *The Lancet Infectious Diseases*, vol. 20, no. 5, pp. e102 – e107, 2020. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1473309920301298

[2] D. Bitar, A. Goubar, and J.-C. Desenclos, "International travels and fever screening during epidemics: a literature review on the effectiveness and potential use of non-contact infrared thermometers," *Eurosurveillance*, vol. 14, no. 6, p. 19115, 2009.

[3] E. Y. Ng, G. Kawb, and W. Chang, "Analysis of IR thermal imager for mass blind fever screening," *Microvascular Research*, vol. 68, no. 2, pp. 104 – 109, 2004. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0026286204000548

[4] E. Ring, H. Mcevoy, A. Jung, J. Zuber, and G. Machin, "New standards for devices used for the measurement of human body temperature," *Journal of medical engineering & technology*, vol. 34, pp. 249–53, 05 2010.

[5] J. B. Mercer and E. F. J. Ring, "Fever screening and infrared thermal imaging: concerns and guidelines," *Thermology International*, vol. 19, no. 3, pp. 67–69, 2009.

[6] V. Blanz and T. Vetter, "A morphable model for the synthesis of 3d faces," in *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*, 1999, pp. 187–194.

[7] B. Amberg, R. Knothe, and T. Vetter, "Expression invariant 3d face recognition with a morphable model," in *2008 8th IEEE International Conference on Automatic Face & Gesture Recognition*. IEEE, 2008, pp. 1–6.

[8] I. Masi, C. Ferrari, A. Del Bimbo, and G. Medioni, "Pose independent face recognition by localizing local binary patterns via deformation components," in *2014 22nd International Conference on Pattern Recognition*. IEEE, 2014, pp. 4477–4482.

[9] T. Neumann, K. Varanasi, S. Wenger, M. Wacker, M. Magnor, and C. Theobalt, "Sparse localized deformation components," *ACM Trans. Graphics*, vol. 32, no. 6, pp. 179:1–179:10, 2013.

[10] L. Galteri, C. Ferrari, G. Lisanti, S. Berretti, and A. Del Bimbo, "Deep 3d morphable model refinement via progressive growing of conditional generative adversarial networks," *Computer Vision and Image Understanding*, vol. 185, pp. 31–42, 2019.

[11] A. Tuan Tran, T. Hassner, I. Masi, E. Paz, Y. Nirkin, and G. Medioni, "Extreme 3d face reconstruction: Seeing through occlusions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 3935–3944.

[12] F. C. Staal, A. J. Ponniah, F. Angullia, C. Ruff, M. J. Koudstaal, and D. Dunaway, "Describing crouzon and pfeiffer syndrome based on principal component analysis," *Journal of Cranio-Maxillofacial Surgery*, vol. 43, no. 4, pp. 528–536, 2015.

[13] X. Jin and X. Tan, "Face alignment in-the-wild: A survey," *Computer Vision and Image Understanding*, vol. 162, pp. 1 – 22, 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1077314217301455

[14] Y. Wu and Q. Ji, "Facial landmark detection: A literature survey," *International Journal of Computer Vision*, vol. 127, no. 2, pp. 115–142, 2019. [Online]. Available: https://doi.org/10.1007/s11263-018-1097-z

[15] G. Hermosilla Vigneau, J. L. Verdugo, G. Farias Castro, F. Pizarro, and E. Vera, "Thermal face recognition under temporal variation conditions," *IEEE Access*, vol. 5, pp. 9663–9672, 2017.

[16] "Notre dame collection x1," 2020. [Online]. Available: https://cvrl.nd.edu/projects/data/

[17] R. S. Ghiass, BendadaHakim, and X. Maldague, "Université Laval face motion and time-lapse video database (UL-FMTV)," in *Proc. of Quantitative InfraRed Thermography Conference (QIRT)*, 2018.

[18] M. Kopaczka, R. Kolk, and D. Merhof, "A fully annotated thermal face database and its application for thermal facial expression recognition," in *Proc. of IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*. IEEE, 2018, pp. 1–6.

[19] 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.

[20] Z. Wu, M. Peng, and T. Chen, "Thermal face recognition using convolutional neural network," in *Proc. of International Conference on Optoelectronics and Image Processing (ICOIP)*. IEEE, 2016, pp. 6–9.

[21] S. Kumar and S. K. Singh, "Occluded thermal face recognition using bag of CNN (*bo*cnn)," *IEEE Signal Processing Letters*, vol. 27, pp. 975–979, 2020.

[22] D. Poster, S. Hu, N. Nasrabadi, and B. Riggan, "An examination of deep-learning based landmark detection methods on thermal face imagery," in *Proc. of the Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2019.

[23] M. Kopaczka, R. Kolk, J. Schock, F. Burkhard, and D. Merhof, "A thermal infrared face database with facial landmarks and emotion labels," *IEEE Transactions on Instrumentation and Measurement*, vol. 68, no. 5, pp. 1389–1401, 2019.

[24] M. Kopaczka, J. Schock, J. Nestler, K. Kielholz, and D. Merhof, "A combined modular system for face detection, head pose estimation, face tracking and emotion recognition in thermal infrared images," in *Proc. of IEEE International Conference on Imaging Systems and Techniques (IST)*, 2018, pp. 1–6.

[25] W. Chu and Y. Liu, "Thermal facial landmark detection by deep multi-task learning," in *Proc. of IEEE International Workshop on Multimedia Signal Processing (MMSP)*, 2019, pp. 1–6.

[26] Z. Cao, G. Hidalgo Martinez, T. Simon, S. Wei, and Y. A. Sheikh, "Openpose: Realtime multi-person 2D pose estimation using part affinity fields," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019.

[27] C. Ferrari, G. Lisanti, S. Berretti, and A. Del Bimbo, "A dictionary learning-based 3D morphable shape model," *IEEE Transactions on Multimedia*, vol. 19, no. 12, pp. 2666–2679, 2017.

[28] R. Vardasca *et al.*, "The influence of angles and distance on assessing inner-canthi of the eye skin temperature," *Thermology international*, vol. 27, no. 4, pp. 130–135, 2017.

[29] M. Kopaczka, T. Jantos, and D. Merhof, "Towards analysis of mental stress using thermal infrared tomography," in *Bildverarbeitung für die Medizin 2018*. Springer, 2018, pp. 157–162.

[30] N. Srivastava, "Using contactless sensors to estimate learning difficulty in digital learning environments," in *Proc. of ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp/ISWC)*, ser. UbiComp/ISWC '19 Adjunct. New York, NY, USA: Association for Computing Machinery, 2019, pp. 399–403. [Online]. Available: https://doi.org/10.1145/3341162.3349312

[31] L.-S. Chan, G. T. Y. Cheung, I. J. Lauder, and C. R. Kumana, "Screening for Fever by Remote-sensing Infrared Thermographic Camera," *Journal of Travel Medicine*, vol. 11, no. 5, pp. 273–279, 03 2006. [Online]. Available: https://doi.org/10.2310/7060.2004.19102

[32] D. D. Pascoe, E. F. Ring, J. B. Mercer, J. Snell, D. Osborn, and J. Hedley-Whyte, "International standards for pandemic screening using infrared thermography," in *Medical Imaging 2010: Biomedical Applications in Molecular, Structural, and Functional Imaging*, R. C. Molthen and J. B. Weaver, Eds., vol. 7626, International Society for Optics and Photonics. SPIE, 2010, pp. 589 – 596. [Online]. Available: https://doi.org/10.1117/12.843836

[33] S. Budzan and R. Wyżgolik, "Face and eyes localization algorithm in thermal images for temperature measurement of the inner canthus of the eyes," *Infrared Physics & Technology*, vol. 60, pp. 225 – 234, 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1350449513001096

[34] J.-W. Lin, M.-H. Lu, and Y.-H. Lin, "A thermal camera based continuous body temperature measurement system," in *Proc. of the International Conference on Computer Vision (ICCV) Workshops*, Oct 2019.

[35] A. Aryal and B. Becerik-Gerber, "Skin temperature extraction using facial landmark detection and thermal imaging for comfort assessment," in *Proc. of ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*, ser. BuildSys '19. New York, NY, USA: Association for Computing Machinery, 2019, pp. 71–80. [Online]. Available: https://doi.org/10.1145/3360322.3360848

[36] S. D. Lin, K. Chen, and W. Chen, "Thermal face recognition based on physiological information," in *Proc. of IEEE International Conference on Image Processing (ICIP)*, 2019, pp. 3497–3501.

[37] C. Ferrari, G. Lisanti, S. Berretti, and A. Del Bimbo, "Dictionary learning based 3D morphable model construction for face recognition with varying expression and pose," in *Proc. of International Conference on 3D Vision*, 2015, pp. 509–517.

[38] S. Katz, A. Tal, and R. Basri, "Direct visibility of point sets," in *Proc. of ACM SIGGRAPH*, 2007, pp. 24–es.

[39] V. Le, J. Brandt, Z. Lin, L. Bourdev, and T. S. Huang, "Interactive facial feature localization," in *Proc. of European Conference on Computer Vision (ECCV)*. Springer, 2012, pp. 679–692.

[40] C. Ferrari, G. Lisanti, S. Berretti, and A. Del Bimbo, "Effective 3d based frontalization for unconstrained face recognition," in *2016 23rd International Conference on Pattern Recognition (ICPR)*. IEEE, 2016, pp. 1047–1052.

[41] C. Qu12, E. Monari, T. Schuchert, and J. Beyerer21, "Adaptive contour fitting for pose-invariant 3d face shape reconstruction," 2015.

[42] A. Bulat and G. Tzimiropoulos, "How far are we from solving the 2d & 3d face alignment problem? (and a dataset of 230,000 3d facial landmarks)," in *International Conference on Computer Vision*, 2017.

[43] X. Yu, W. K. Chua, L. Dong, K. E. Hoe, and L. Li, "Head pose estimation in thermal images for human and robot interaction," in *2010 The 2nd International Conference on Industrial Mechatronics and Automation*, vol. 2. IEEE, 2010, pp. 698–701.

[44] M. Kopaczka, J. Schock, J. Nestler, K. Kielholz, and D. Merhof, "A combined modular system for face detection, head pose estimation, face tracking and emotion recognition in thermal infrared images," in *2018 IEEE International Conference on Imaging Systems and Techniques (IST)*. IEEE, 2018, pp. 1–6.

[45] T.-Y. Yang, Y.-T. Chen, Y.-Y. Lin, and Y.-Y. Chuang, "Fsa-net: Learning fine-grained structure aggregation for head pose estimation from a single image," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 1087–1096.