Inner Eye Canthus Localization for Human Body Temperature Screening

Claudio Ferrari, Lorenzo Berlincioni, Marco Bertini, Alberto Del Bimbo

Media Integration and Communication Center (MICC)
University of Florence

Motivation

Infrared thermography has a huge practical importance in non intrusively detecting body fever, common precursor of many diseases like H1N1, SARS, MERS and COVID-19.

Its relevance as a tool for preventing the diffusion of diseases requires a proper implementation, and to adhere to international safety ISO standards [*]:

1. Evaluate the temperature in correspondence of the inner eye canthus.
   - Warmest face region, most invariant to environmental conditions and exertion.

2. Perform measurement on a region rather than a single point to account for sensor noise.

3. Subject should face the camera.
   - Slight head rotations lead to measurements errors up to 1-2 degrees.

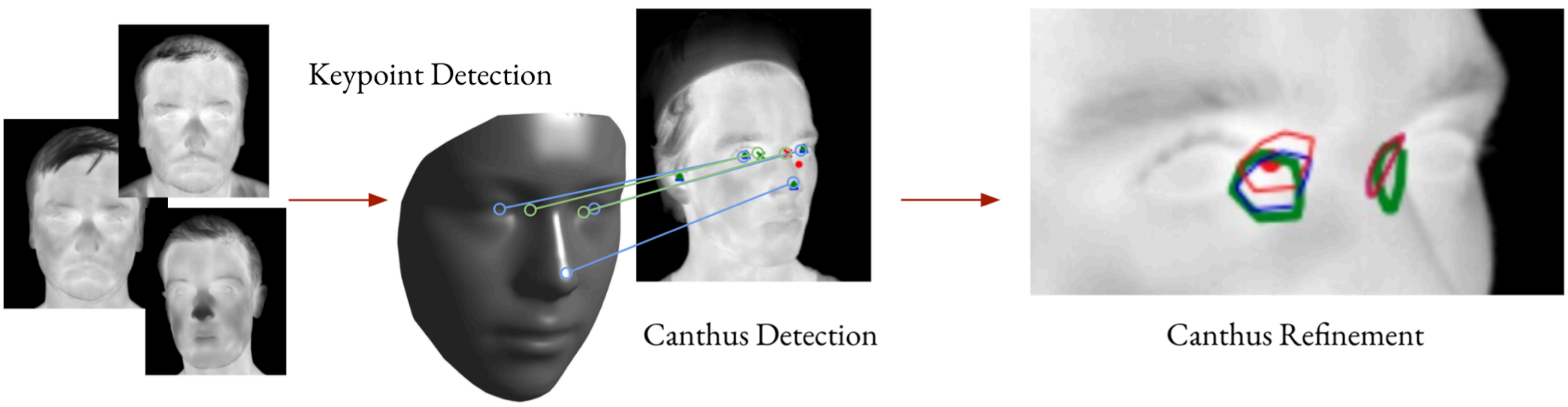
Accurately localizing the inner eye canthus region is thus of outmost importance!

[*] https://www.iso.org/standard/69347.html

Proposed Method

The proposed algorithm consists of three main steps:

1. Use OpenPose [*] to coarsely localize 5 facial keypoints: center of the eyes, nosetip and ears.
2. Exploit a 3D Morphable Face Model (3DMM) to estimate the head pose and a 3D-2D projection;
    - Project the 3DMM onto the image to localize the inner canthus.
3. Refine the localization of the inner canthus by a warmest point search strategy.



The only manual intervention required is to label the 5 keypoints on the 3D model.

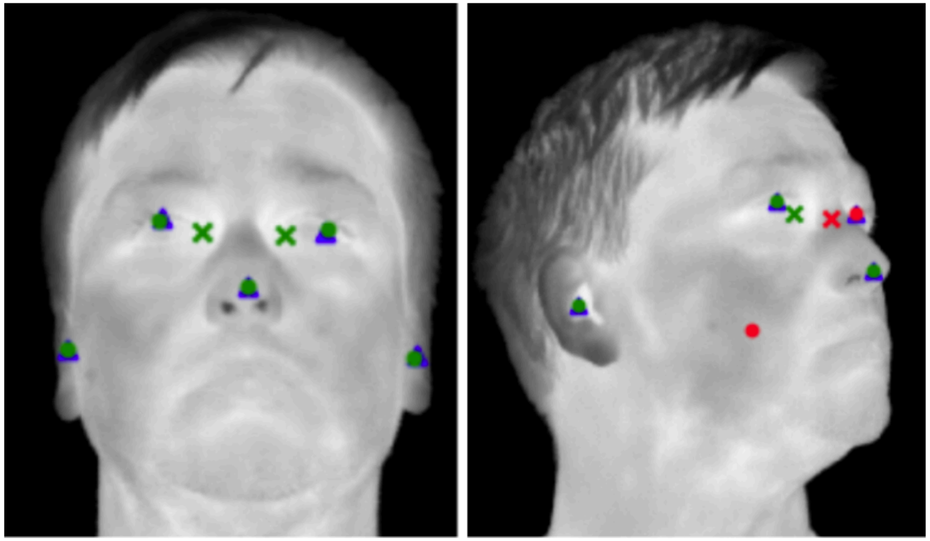
[*] Cao, Zhe, et al. "OpenPose: realtime multi-person 2D pose estimation using Part Affinity Fields." *IEEE transactions on pattern analysis and machine intelligence* (2019).

Head Pose Estimation

- Given the detected 2D keypoints $l_{op} \in \mathbb{R}^{2\times5}$ and the corresponding 3D points $\mathrm{L}_{\mathrm{op}} \in \mathbb{R}^{3\times5}$ we estimate the head pose using an orthographic camera model:

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

- The camera matrix $A \in \mathbb{R}^{2\times3}$ and translation $t \in \mathbb{R}^2$ are estimated solving a least squares problem.
- QR decomposition is applied to A to get 3D rotation and scale matrices $\mathrm{R} \in \mathbb{R}^{3\times3}$ and $S \in \mathbb{R}^{2\times3}$



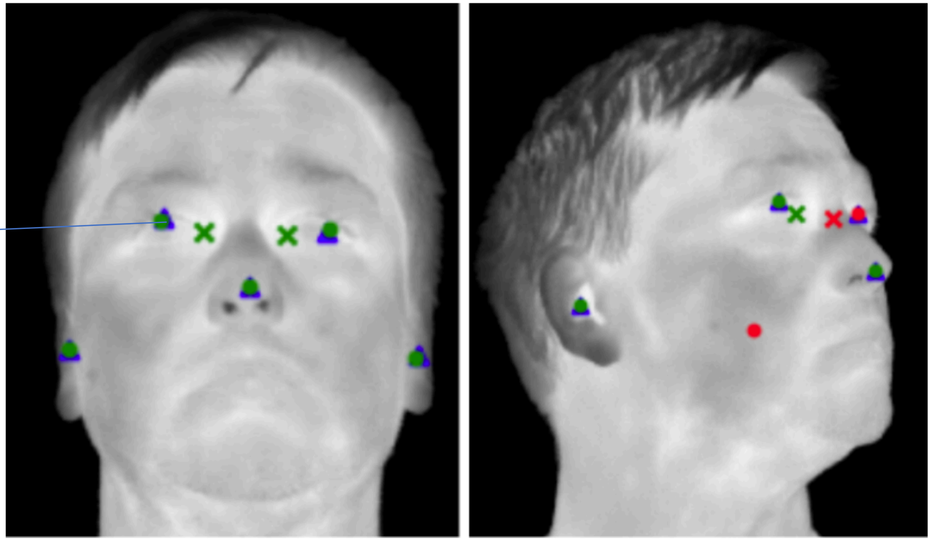
Head Pose Estimation

- Given the detected 2D keypoints $l_{op} \in \mathbb{R}^{2\times5}$ and the corresponding 3D points $\mathrm{L_{op}} \in \mathbb{R}^{3\times5}$ we estimate the head pose using an orthographic camera model:

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

- The camera matrix $A \in \mathbb{R}^{2\times3}$ and translation $t \in \mathbb{R}^{2}$ are estimated solving a least squares problem.
- QR decomposition is applied to A to get 3D rotation and scale matrices $\mathrm{R} \in \mathbb{R}^{3\times3}$ and $S \in \mathbb{R}^{2\times3}$

Blue: OpenPose Keypoints
Green: Projected 3D Keypoints



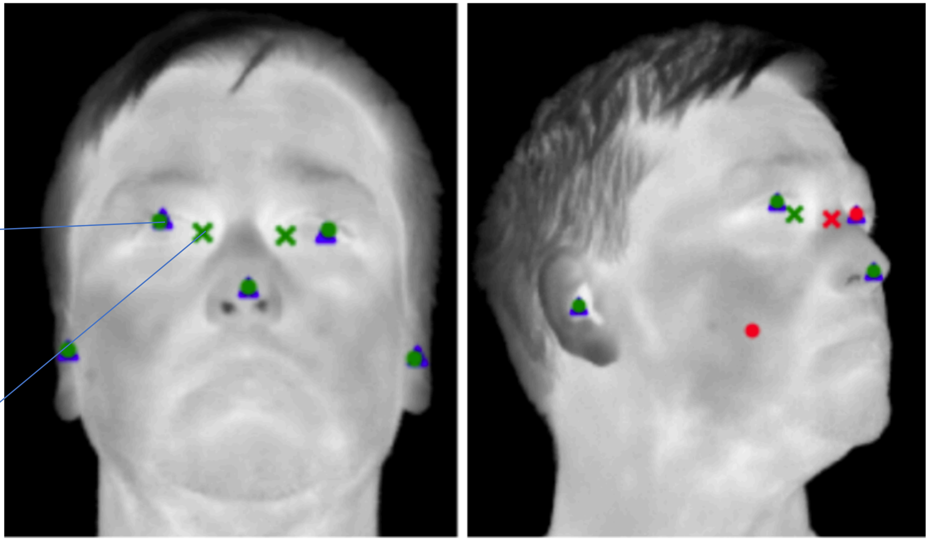
Head Pose Estimation

- Given the detected 2D keypoints $l_{op} \in \mathbb{R}^{2\times5}$ and the corresponding 3D points $\mathrm{L_{op}} \in \mathbb{R}^{3\times5}$ we estimate the head pose using an orthographic camera model:

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

- The camera matrix $A \in \mathbb{R}^{2\times3}$ and translation $t \in \mathbb{R}^2$ are estimated solving a least squares problem.
- QR decomposition is applied to A to get 3D rotation and scale matrices $\mathrm{R} \in \mathbb{R}^{3\times3}$ and $S \in \mathbb{R}^{2\times3}$

Blue: OpenPose Keypoints
Green: Projected 3D Keypoints



Green Cross: Projected inner Eyes canthus from the 3D model.

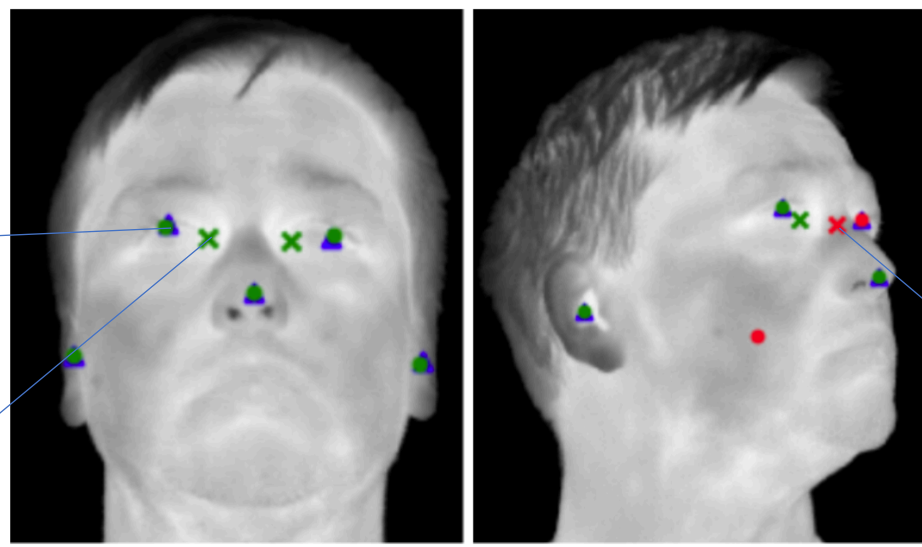
Head Pose Estimation

- Given the detected 2D keypoints $l_{op} \in \mathbb{R}^{2\times5}$ and the corresponding 3D points $\mathrm{L_{op}} \in \mathbb{R}^{3\times5}$ we estimate the head pose using an orthographic camera model:

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

- The camera matrix $A \in \mathbb{R}^{2\times3}$ and translation $t \in \mathbb{R}^2$ are estimated solving a least squares problem.
- QR decomposition is applied to A to get 3D rotation and scale matrices $\mathrm{R} \in \mathbb{R}^{3\times3}$ and $S \in \mathbb{R}^{2\times3}$

Blue: OpenPose Keypoints
Green: Projected 3D Keypoints



Red Cross/Circles: Detect self occlusions by estimating a visibility mask give the viewpoint

Green Cross: Projected inner Eyes canthus from the 3D model.

Eye Canthus Region Refinement

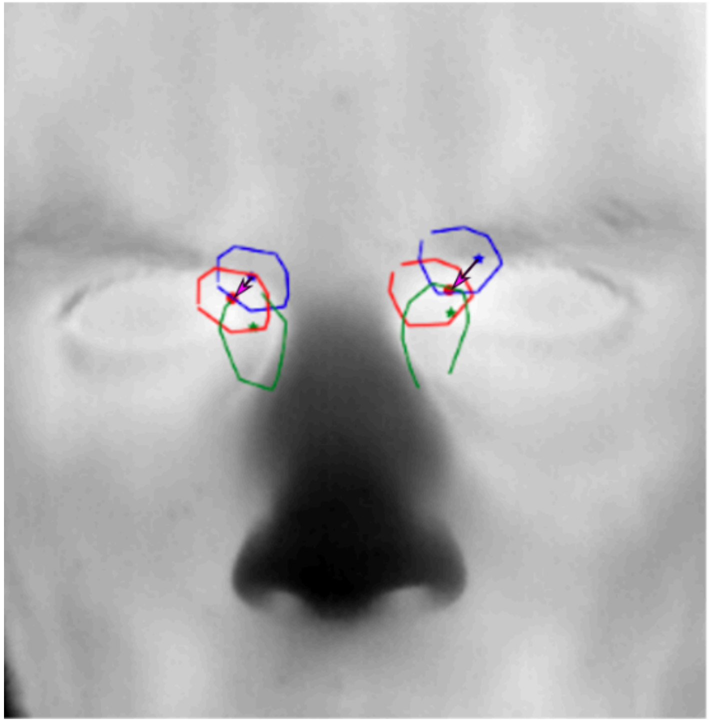
- The estimated projection heavily relies on the accuracy of OpenPose.
- The canthus location is refined by searching for the hottest point within a local region.

Strategy

- Define the convex hull of the k-ring of the projected canthus points;
- Find the hottest point *i.e.* brightest pixel, within the region;
- Estimate the new canthus region.

Advantages:

- Correct possible slight detection errors;
- The measured temperature represents a local upper bound, so dangerous false negatives can be avoided;
- Independent from possible errors in the manual annotation.



Blue: Estimated canthus
Green: Ground-truth location
Red: refined localization of the canthus

Experimental Results

We evaluated the approach on the Thermal FaceDB dataset [*]:

- 2935 thermal frames of 90 subjects annotated with 68 facial landmarks;
- Variations in pose and expressions;
- Additional annotations of head poses, and occlusions have been added.

Canthus detection accuracy

		FAN [16]	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	-

Head Pose Estimation

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

$$(e_\alpha,\ e_\beta,\ e_\gamma) = (8.98\ ,\ 7.16\ ,\ 6.62\ )$$

Execution time (FPS/Sec)

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

[*] M. Kopaczka, R. Kolk, and D. Merhof, “A fully annotated thermal face database and its application for thermal facial expression recognition,” *I2MTC*.

Experimental Results

We evaluated the approach on the Thermal FaceDB dataset [*]:

- 2935 thermal frames of 90 subjects annotated with 68 facial landmarks;
- Variations in pose and expressions;
- Additional annotations of head poses, and occlusions have been added.

Canthus detection accuracy

		FAN [16]	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	-

Head Pose Estimation

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

$$(e_\alpha,\ e_\beta,\ e_\gamma) = (8.98\ ,\ 7.16\ ,\ 6.62\ )$$

Execution time (FPS/Sec)

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

[*] M. Kopaczka, R. Kolk, and D. Merhof, “A fully annotated thermal face database and its application for thermal facial expression recognition,” *I2MTC*.

Thank You!

Claudio Ferrari, Lorenzo Berlincioni, Marco Bertini, Alberto Del Bimbo

Media Integration and Communication Center (MICC)
University of Florence