

Driver Monitoring Systems in automated interactions: A real-time, thermographic-based algorithm

Saifeddine Aloui¹, Raphaël Morvillier¹, Christophe Prat¹, Jaka Sodnik², Carolina Diaz-Piedra³, Francesco Angioi³ and Leandro L. Di Stasi³

¹ Univ. Grenoble Alpes, CEA, Leti, F-38000 Grenoble, France

² University of Ljubljana. Faculty of Electrical Engineering, Tržaška c. 25, 1000 Ljubljana, Slovenia

³ Mind, Brain, and Behavior Research Center – CIMCYC, University of Granada, Campus de Cartuja s/n, Granada 18011, Spain

Abstract

Due to the progressive shift of responsibility from the driver to the vehicle itself in automated vehicle technologies, driver-centered innovations represent a key point for its advance. The so-called Driver Monitoring Systems (DMS) are therefore increasingly gaining importance in this context. One of the main aims of DMS is to estimate the driver's arousal levels in order to infer their cognitive state and capabilities. Even though the scientific literature is riddled with useful psychophysiological indices to estimate arousal levels [1], nowadays, arousal estimation is based on broad, mostly blink/gaze-related, indices. The reason is that actual implementation of reliable sensors in a feasible system able to collect, analyze, and interpret measurements in real-life conditions is still an open challenge. One of the alternatives to signal different cognitive states is facial skin temperature [2][3]. Infrared sensors that monitor heat loss have been shown useful to track facial skin temperature that indicate arousal modulations while driving [2][3]. Such intensive, laborious work to extract and analyze temperature changes in some facial landmarks is not reasonable in real-life applications [2]. Here, we present the preliminary results obtained with a new software able to track, in real-time, drivers' facial-skin temperature changes.

Keywords

Driver state; Workload; Facial thermography; Real-time algorithm; Automated vehicle; Sensing and real-time information

1. Introduction

Due to the progressive shift of responsibility from the driver to the vehicle itself in automated vehicle technologies, driver-centered innovations represent a key point for its advance. The so-called Driver Monitoring Systems (DMS) are therefore increasingly gaining importance in this context. One of the main aims of DMS is to estimate the driver's arousal levels in order to infer their cognitive state and capabilities. Even though the scientific literature is riddled with useful psychophysiological indices to estimate arousal levels [1], nowadays, arousal estimation is based on broad, mostly blink/gaze-related, indices. The reason is that actual implementation of reliable sensors in a feasible system able to collect, analyze, and interpret measurements in real-life conditions is still an open challenge. One of the alternatives to signal different cognitive states is facial skin temperature [2,3]. Infrared sensors that monitor heat loss have been shown useful to track facial skin temperature that indicate arousal modulations while driving. Such intensive, laborious work to extract and analyze temperature changes in some facial landmarks is not reasonable in real-life applications [2]. Face landmarks extraction using color images has become of common use [4] thanks to several libraries (e.g., Google's MediaPipe

HCISI 2022, November 29, 2022, Ljubljana, Slovenia

EMAIL: saifeddine.aloui@cea.fr (A. 1); raphael.morvillier@cea.fr (A. 2); christophe.prat@cea.fr (A. 3); jsodnik@nervtech.com (A. 4); dipie@ugr.es (A. 5); frangioi@ugr.es (A. 6); distasi@ugr.es (A. 7)

ORCID: 0000-0001-7020-4461 (A.1); 0000-0003-2868-331X (A.2); 0000-0001-8074-885X (A.3); 0000-0002-8915-9493 (A.4); 0000-0002-8168-2546 (A. 5); 0000-0001-8231-5580 (A. 6); 0000-0001-6763-6546 (A. 7)



© 2020 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

library [5]). However, when applied to thermographic images, these libraries produce unsatisfactory results: the face is either not detected or the landmarks are not correctly aligned with the real face. Therefore, two main methods have been developed to perform landmark detection on thermographic images. The first method is to develop a dedicated system trained on annotated thermographic images [see 6]. This approach is still limited due to the lack of large thermographic databases. For example, Kopaczka and colleagues used a database containing 2,935 images [7]. A database of this kind would not be useful for our data. In the present study, the drivers had to wear transparent face masks due to the COVID19 pandemic. This made it harder to apply landmark detection on thermographic images where the mask was visible. Indeed, although the masks were transparent to visible light, they were not in the wavelength used to measure the temperature, therefore hiding part of the driver's face. The second method uses an additional color camera to detect the facial landmarks and transfers them on the thermographic image (this process of aligning images from different sources is often referred to as "image registration"). In previous studies, authors detected the edges in both color and thermal images and match them to align the images [8, 9]. A simpler method is described in another work [10], based on an initial optical calibration between the two cameras. Goulart and colleagues used the same principle and add a post-processing step to enhance the transferred landmark position, based on a trained expert manual annotation [11]. Here, we present the preliminary results obtained with a system based on this second method, able to track drivers' facial-skin temperature changes automatically after an initial calibration. It is a first step towards a fully automatic system, which could run in real-time in future vehicles. We present the principle of the system and analyze its performance. In a future work, we intend to show the usefulness of extracting the face temperature in an automated driving condition.

2. Material and methods

2.1. Instruments

We used a sensorized driving simulator (Nervtech™ solution, see Figure 1) running a SCANer studio software (AVSimulation, v.DT2.5). Participants' facial skin temperature was constantly monitored with a thermographic camera (FLIR A325sc, with a resolution of 320×240 , a NETD $< 50\text{mK}$ and an accuracy of $\pm 2^\circ\text{C}$ or $\pm 2\%$ of reading) synchronized with a color camera (infrared color camera, Intel® Realsense).

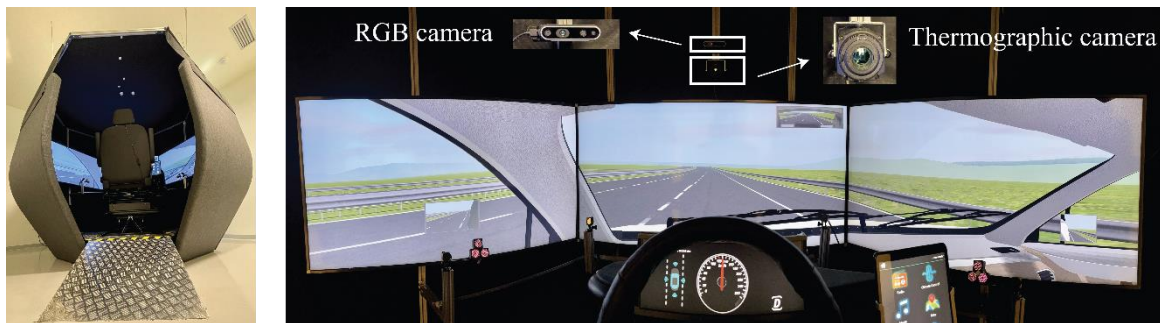


Figure 1: The driving simulator employed in the study. Left, the simulator and its dome; right, the interior of the dome with the thermographic and color cameras on the top of the main screen.

2.2. Face temperature extraction algorithm

To extract participants' facial-skin temperature at specific locations, we developed an algorithm (Figure 2) able to identify two facial landmarks (Points of Interest [POIs]), the tip of the nose and the forehead, as well as the background in a thermographic image. The solution was based on a dual camera setup (i.e., color camera and thermographic camera), with a spatial correspondence between the two.

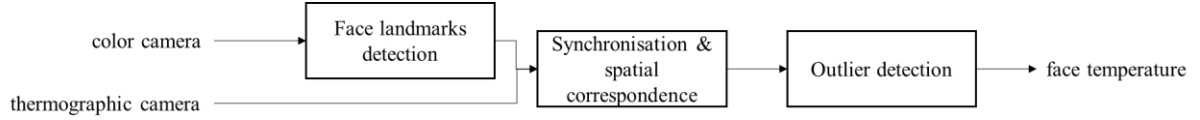


Figure 2: Architecture of the temperature extraction algorithm

Specifically, the color camera output allows the extraction of the POIs using conventional facial landmarks extraction tools. Here, we used MediaPipe [5], the state-of-the-art landmark detection library. Once the POIs were detected, their positions were fitted into the thermographic camera output, using a geometric transformation [12]. The algorithm uses a 3×3 transformation matrix ($T^{color \rightarrow thermal}$) to convert each POI position from the color camera spatial output to the thermographic spatial output. Each POI (landmark, l_i) is defined by its coordinates in the color camera space (l_i^{color}). As detailed below, the coordinates in the thermographic camera space ($l_i^{thermal}$) are obtained by multiplying l_i^{color} by the transformation matrix $T^{color \rightarrow thermal}$ (1). This transformation matrix is the multiplication of three matrices (2). The first describes a translation with coordinates $[t_x, t_y]$ (3), the second describes a rotation around the center of the screen with angle θ (4) and the third describes a scaling with parameters $[s_x, s_y]$ (5). Once the positions of the landmarks in the thermographic image space were found, the POIs temperature values were read in the image. Finally, we multiplied the result by the skin emissivity (0.98) to obtain the skin temperature.

$$l_i^{thermal} = T^{color \rightarrow thermal} l_i^{color} \quad (1)$$

$$T^{RGB \rightarrow thermal} = T^{trans} \times T^{rot} \times T^{scaling} \quad (2)$$

$$T^{trans} = \begin{bmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \\ 0 & 0 & 1 \end{bmatrix} \quad (3)$$

$$T^{rot} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) & 0 \\ \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (4)$$

$$T^{scaling} = \begin{bmatrix} s_x & 0 & 0 \\ 0 & s_y & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (5)$$

2.3. Calibration

The described system first needed to be calibrated to determine the parameters of the transformation matrix $T^{color \rightarrow thermal}$: t_x , t_y , θ , s_x and s_y . Filippini and colleagues used a similar set-up and performed the calibration using a custom checkerboard [10], a method we found to be less precise in our situation. We therefore developed a dedicated calibration software. It allows an operator to visualize simultaneously the color and the thermographic camera outputs, as shown in Figure 3.

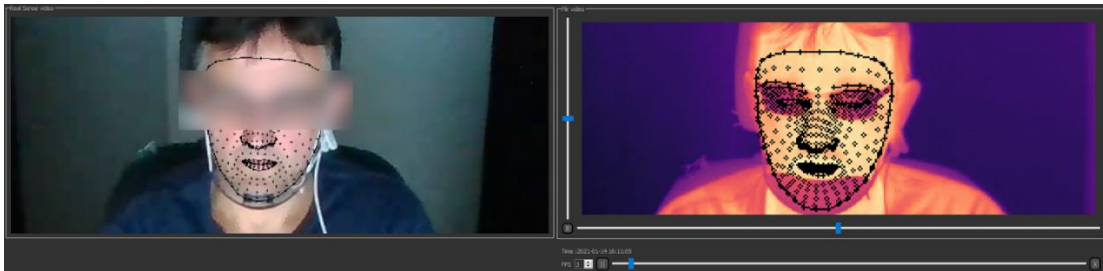


Figure 3: Calibration software interface. On the left, the color image with the landmarks detected thanks to MediaPipe [5]. On the right, the thermographic image with the corresponding landmarks that the operator has to translate, rotate, and scale to correspond to the driver's face.

On the color image, the operator can inspect the landmark detection performed by MediaPipe. A thermographic image shows if these landmarks are transferred correctly. If the result is not satisfying, additional translations, rotations and scaling of the landmarks “mask” can be done manually with the mouse. These transformations are recorded by the calibration software to compute the matrix $T^{color \rightarrow thermal}$. The calibration software finally saves the conversion parameters in a dedicated file which is used by the extraction software to automatically detect the POIs on the thermographic output. In our experiment, we repeated the calibration procedure for each driver to compensate for slight differences in the positions of the cameras and head among different drivers.

2.4. Experimental design

To test our algorithm, we designed a 2 (traffic vs. low-traffic *Traffic density*) \times 2 (automated [ADL4] vs. manual [MD] *Driving modality*) within-participants experiment. Thirty-five expert drivers (mean age = 41.61 years, standard deviation = 6.26 years) drove along two virtual scenarios (\sim 20 minutes [min] each) with varying traffic density. In both scenarios, the participants performed 10 min in MD and 10 min in ADL4. The order of the traffic density and driving mode was randomly balanced across drivers. During ADL4, they were instructed to supervise the system. We expect the arousal level of the drivers to be modulated by these conditions, as the manual driving mode and the high traffic condition are more demanding than the autonomous one and the low traffic condition respectively.

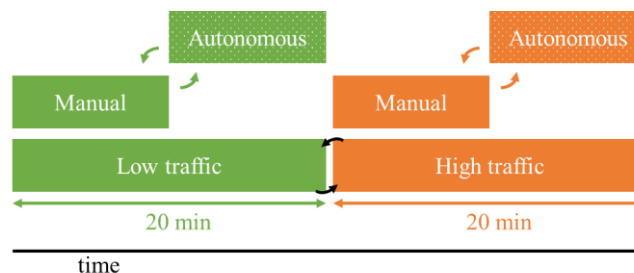


Figure 4: Experimental design. Each participant performed the tasks as it is illustrated above. The arrows indicate that traffic conditions and the driving mode were randomized across participants.

2.5. Validation method

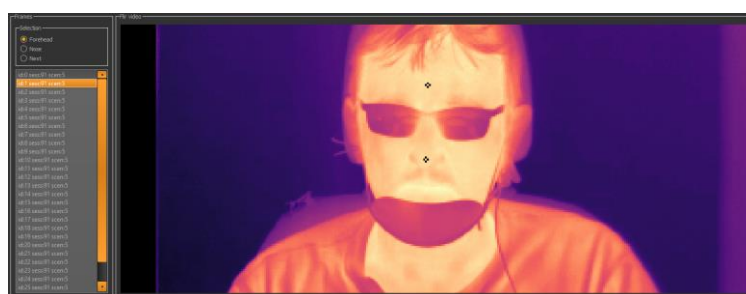


Figure 5: Annotation software interface. Left side: the annotator selects the POI and the frames to be annotated. Right side: the annotator points at the location of the POI on the thermographic image (in this example, the forehead and the nose tip are already annotated).

In order to validate the proposed algorithm, we selected randomly one of the two 20-min recordings (high or low traffic) for each driver. Then, we extracted one pair of color and thermographic images each 20 seconds. We obtained 65 images per driver and 2,340 in total. We then developed an annotation software to manually extract the temperature on these images. For each image, we pointed at two landmarks: the driver’s forehead and the driver’s nose tip. Four trained annotators performed the same procedure on the 2,340 images. Then, we computed on each image the mean of the four annotated

positions of the nose tip and the forehead to establish the reference location of the nose tip and the forehead. Finally, we extracted the temperature at these locations to define the reference temperature.

2.6. Statistical study on the obtained data

After removing the drivers on who the algorithm performed worse (see section 3.1.2), we used the algorithm described in Section 2.2 to extract the face temperature of the remaining drivers ($n = 28$). In order to obtain more measurement points, the algorithm at this stage ran at a higher frequency compared to the validation phase: one each 2 sec instead of one each 20 sec. We were therefore able to remove extreme values (lower than 25°C and higher than 37°C) as well as the outliers by applying a moving median thresholding procedure. We finally took the mean of the remaining points on each of the four segments: High traffic – Manual driving, High traffic – Automated driving, Low traffic – Manual driving, Low traffic – Automated driving. This gave us four data points per driver that we later used in our statistical analysis.

3. Results

3.1. Validation of the algorithm

We first analyzed the algorithm performance in terms of position error in the thermographic image, measured in pixels. We computed the position error between the algorithm's output and the mean values provided by the four annotators (see 2.5). We also compared the position error of the mean position error of each annotator with respect to the overall mean value. Then, we analyzed the consequences of the algorithm position error in terms of temperature error, measured in degrees Celsius ($^{\circ}\text{C}$).

3.1.1. Position error

In Figure 6, we present the errors' distributions of the algorithm and the annotators. As a reference, in our setup the nose tip measures approximately 10×10 pixels. When pointing at the nose, the algorithm performed worse than the annotators with respect to the mean of the annotators. The two main causes for high mismatches were landmarks estimation errors of MediaPipe and spatial correspondence errors due to head movements (head turning or bending). Surprisingly, the algorithm outperformed slightly the annotators on the forehead with respect to the mean of the annotators. Our interpretation is that for a human, it could be hard to define a precise location on a large area with no points of reference such as the forehead.

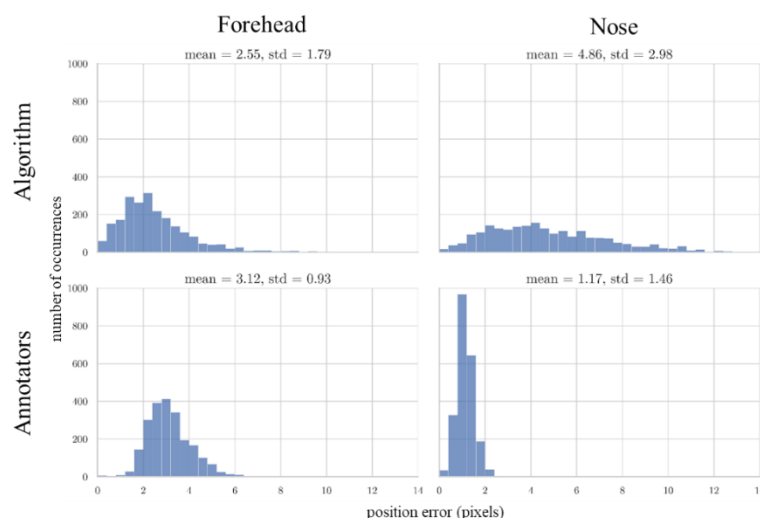


Figure 6: Position error (a 2D distance in pixels) distribution for the forehead and the nose. At the top, we compared the algorithm to the mean of the annotators. At the bottom, we compared each annotator to the mean of the annotators.

3.1.2. Temperature error

Figure 7 shows the errors of the final temperature values computed by the algorithm. On the forehead, the temperature gradient was low, so the temperature error resulting from the position error was small. On the nose, however, the temperature gradient was higher, so the temperature error was also much higher compared to the forehead. Interestingly, the temperatures computed at the positions annotated by one annotator are consistently smaller than the temperatures computed at the mean of the annotated positions. This is because the face temperature exhibits a local peak on the nose and one individual annotator is further from this peak than the mean position of the four annotators.

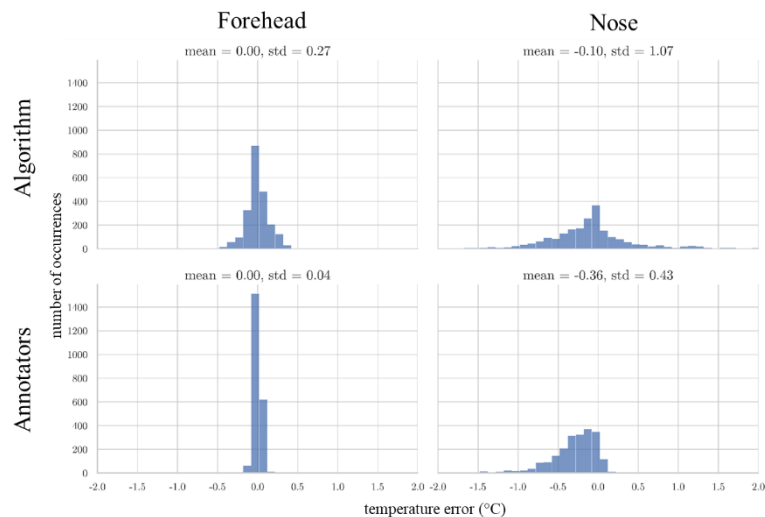


Figure 7: Temperature error (in °C) distribution for the forehead and the nose. At the top, we compared the algorithm to the mean of the annotators. At the bottom, we compared each annotator to the mean of the annotators.

Looking at Figure 8, we see that the mean absolute error of the nose temperature highly depends on the driver (it goes up to 1.6 °C for some drivers). For the statistical study, we excluded the 6 participants with an absolute error higher than 0.8 °C.

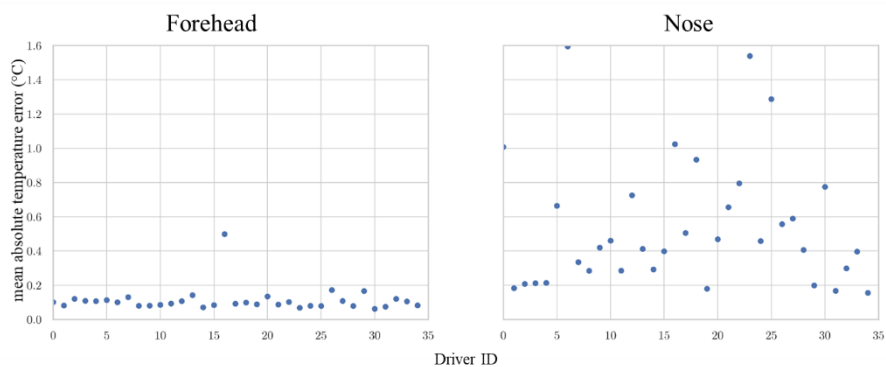


Figure 8: Algorithm mean absolute temperature error, for each driver (in °C).

4. Conclusion and future works

The present work describes the first results obtained with an algorithm for tracking a driver's facial skin temperature during driving interactions. The algorithm consistently and effectively tracked participants' facial-skin temperature without interfering with their driving tasks. We have analyzed the position and temperature errors and for some drivers, tracking the nose tip temperature remains a challenge. Future systems should improve both the initial landmarks detection and the landmark

transfer. The later could be achieved by measuring the distance between the cameras and the driver's face like previous studies [10] or considering the face as a 3D shape. Also, a calibration-less process should be developed to be implemented in a real car. More analysis should be conducted before publishing the results of a statistical study based on this work.

5. Acknowledgements

This study was funded by the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875597 - HADRIAN (Holistic Approach for Driver Role Integration and Automation Allocation for European Mobility Needs) project. This document reflects only the authors' view, the European Climate, Infrastructure and Environment Executive Agency (CINEA) is not responsible for any use that may be made of the information it contains. We thank Leila Maboudi (Polytechnic University of Turin, Italy) for her comments and assistance in language edition.

6. References

- [1] L.L. Di Stasi, E. Gianfranchi, C. Diaz-Piedra, Hand-skin temperature response to driving fatigue: an exploratory study, in: Krömker, H. (Eds), *HCI in Mobility, Transport, and Automotive Systems. Driving Behavior, Urban and Smart Mobility, HCII 2020*, vol 12213 of *Lecture Notes in Computer Science*, Springer, Cham, 2020, pp. 3–14. doi:10.1007/978-3-030-50537-0_1.
- [2] C. Diaz-Piedra, E. Gomez-Milan, L.L. Di Stasi. Nasal skin temperature reveals changes in arousal levels due to time on task: An experimental thermal infrared imaging study, *Applied Ergonomics* 81 (2019). doi: 10.1016/j.apergo.2019.06.001.
- [3] Panasonic Corporation, Panasonic develops drowsiness-control technology by detecting and predicting driver's level of drowsiness, 2017. URL: <https://news.panasonic.com/global/press/data/2017/07/en170727-3/en170727-3.html>
- [4] M. Bodini. A review of facial landmark extraction in 2D images and videos using deep learning, *big data and cognitive computing* 3, 14 (2019). doi: 10.3390/bdcc3010014.
- [5] C. Lugaresi, J. Tang, H. Nash, C. McClanahan, E. Uboweja, M. Hays, F. Zhang, C.-L. Chang, M.G. Yong, J. Lee, W.-T. Chang, W. Hua, M. Georg, M. Grundmann, 2019. MediaPipe: A framework for building perception pipelines. arXiv:1906.08172.
- [6] W.-T. Chu, Y.-H. Liu. Thermal facial landmark detection by deep multi-task learning, in 2019 IEEE 21st International Workshop on Multimedia Signal Processing (MMSP), 2019, pp. 1–6. doi:10.1109/MMSP.2019.8901710.
- [7] M. Kopaczka, R. Kolk, D. Merhof. A fully annotated thermal face database and its application for thermal facial expression recognition, in 2018 IEEE International Instrumentation and Measurement Technology Conference (I2MTC), 2018, pp. 1–6. doi: 10.1109/I2MTC.2018.8409768.
- [8] H. Yoshikawa, A. Uchiyama, T. Higashino. ThermalWrist: Smartphone thermal camera correction using a wristband sensor †, *Sensors* 19 (2019). doi: 10.3390/s19183826.
- [9] L. Sun, Z. Zheng. Thermal-to-visible face alignment on edge map, *IEEE Access* 5 (2017) 11215–11227. doi: 10.1109/ACCESS.2017.2712159.
- [10] C. Filippini, E. Spadolini, D. Cardone, D. Bianchi, M. Preziuso, C. Sciaretta, V. Del Cimmuto, D. Lisciani, A. Merla. Facilitating the child–robot interaction by endowing the robot with the capability of understanding the child engagement: The case of mio amico robot, *International Journal of Social Robotics* 13 (2019) 677-689. doi: 10.1007/s12369-020-00661-w.
- [11] C. Goulart, C. Valadão, D. Delisle-Rodriguez, D. Funayama, A. Favarato, G. Baldo, V. Binotte, E. Caldeira, T. Bastos-Filho. Visual and thermal image processing for facial specific landmark detection to infer emotions in a child-robot interaction, *Sensors (Basel)* 19 (2019). doi: 10.3390/s19132844.
- [12] R. Artzy, *Linear Geometry*, Dover Publications, New York, NY, 1993.