

Proceedings

Thermal Imaging Based Affective Computing for Educational Robot [†]

Chiara Filippini ^{1,*}, Edoardo Spadolini ², Daniela Cardone ¹ and Arcangelo Merla ^{1,2}

¹ Department of Neurosciences, Imaging and Clinical Sciences, University G. d'Annunzio of Chieti-Pescara, Pescara 66100, Italy; d.cardone@unich.it (D.C.); arcangelo.merla@unich.it (A.M.)

² Next2U! Thermal Imaging solutions s.r.l. Pescara 65127 Italy; e.spadolini@next2u-solutions.com

* Correspondence: chiara.filippini@unich.it; Tel.: +39-08713556954

† Presented at the 15th International Workshop on Advanced Infrared Technology and Applications (AITA 2019), Florence, Italy, 17–19 September 2019.

Published: 23 September 2019

Abstract: Over the recent years, Social Robots (SRs) have become more and more prominent in everyday human lives. The main goal of a SR is to interact and communicate with human by following social behaviors and affective interaction. However, they still encounter significant limitations in pursuing a natural interaction, mainly due to their hard task of recognizing and understanding human emotions thus ensuring an appropriate response. The aim of this study was to enrich the SR with affective computing capability and real time assessment of the interlocutor's psychophysiological state, by means of computational psychophysiology based on thermal infrared imaging.

Keywords: affective computing; human-machine interaction; infrared imaging; social robot; computational psychophysiology

1. Introduction

Over the recent years, an increasing number of studies have confirmed the promise of Social Robots (SRs) in many applications ranging from education, health, entertainment and communication [1]. Focusing on the application with infants, SRs are intended to create close and effective interaction with the children, helping them to improve their learning capability [2]. It has been demonstrated that robots that exhibit appropriate emotional responses motivate the user to produce higher quality training data compared to users that interacted with robot with inappropriate or apathetic emotional responses [3]. Therefore, in this study, a novel technology, based on functional infrared thermal imaging (fIRI), was introduced to ensure a socially contingent interaction between children and robots by allowing the artificial agents to perceive the psychophysiological and emotional states of the child and, ideally—choose an appropriate support strategy based on it. fIRI allows contact-less and non-invasive recording of the cutaneous temperature through the measurement of the spontaneous thermal irradiation of the body [4]. By recording the dynamics of the facial cutaneous temperature, it is possible to assess autonomic nervous system activity and infer the subject's psychophysiological or emotional state [5]. In this regard, an automatic recording of thermal IR imaging data and real-time processing is required. Since until now, real-time processing in realistic scenarios has been conducted by employing high-end thermal IR cameras [6–8], the main challenge that has been addressed in this study was the development of a feasible solution for commercial social robots, integrating consumer market technology and low-cost Original Equipment Manufacturer (OEM) based components. The solution, here proposed, consisted of a Computational Psychophysiological Module (CPM) able to assess the temperature variation in specific Region Of Interest (ROI) located on the children's face, to discern the psychophysiological indicators of the

sympathetic/parasympathetic system's activation and to make it available for real time classification of three macro-levels of their emotional engagement: positive, neutral and negative emotional engagement.

2. Material and Methods

2.1. Participants

The experimental session involved 17 children, aged from 4 to 5 years old. Parents were widely informed about the protocol and the main goal of the study. Informed consent form was signed before the experimental trials began.

2.2. Materials and Data Acquisition

The SR utilized in this study was "Mio Amico" robot, produced by ©Liscianigiochi. A mobile thermal imaging solution, the FLIR ONE (2nd gen), dimensions of $11.8 \times 12.7 \times 7.22$ mm, was installed on the head of the robot. The FLIR ONE's thermal imager is a Long Wavelength Infrared (LWIR) microbolometric camera with a resolution of 160×120 pixels for a horizontal FoV of 55° , a Noise Equivalent Temperature Difference (NETD) of 100 mK, and a radiometric accuracy of $+5^\circ\text{C}/+5\%$; in addition to the thermal camera, the FLIR ONE includes a visible light color camera, with a resolution of 640×480 . The two sensors are held in close proximity, vertically aligned, and frames are captured at the same time for both sensors, at a frame rate of about 9 Hz.

2.3. Procedure

The experimental protocol consisted of an "event related" paradigm. In detail, the events lasted 30 s and specifically they were: i) robot telling a fairy tale, ii) robot singing a song. At the end of each event, the SR asked: "Did you like the fairy tale/song, do you want to listen to another one? The response was used as indicator of the child's level of engagement. The type of the next event was chosen depending on the child's response. Each experimental session consisted of 6 events.



Figure 1. Two examples of robot and child interaction under the supervision of an adult.

2.4. Computational Psychophysiology Module—CPM

The CPM was dedicated to the tracking, extraction and analysis of the child's psychophysiological state. The described approach relied on the visible spectrum camera to provide detection and localization of facial features, which were then mapped onto the thermal image coordinate space for the purpose of ROI localization and signal extraction over time. The coordinate mapping between the two cameras relied on an optical calibration procedure, which allowed the calculation of both the geometric relationship (rotation and translation) between the visible and thermal cameras, and the intrinsic parameters of each camera separately (focal length, coordinates of the principal point and coefficients for the lens distortion model). To estimate these parameters, freely usable programs and libraries designed to work with images in the visible spectrum were used [9]. For data recording and streaming, a Raspberry Pi (model 3B) was used, with customized software made to interface with the FLIR ONE's data protocol and to allow the control of the recordings over local WiFi. All the raw data was kept for post-process analysis.

Thermal Data Extraction and Classification

The signal processing techniques used for extraction and analysis of thermal signals were chosen to avoid both excessive delays and high computational load on the system.

The signal extraction pipeline consisted of the following processes:

- Face detection algorithm, applied on the visible image. In detail, the frontal face detector is based on histogram of oriented gradients (HOG) features and a linear SVM classifier [10]. Faces that appeared rotated off-axis were specifically excluded, to preserve the quality of the signals extracted in the later steps.
- Facial landmarks calculation, using an implementation of One Millisecond Face Alignment with an Ensemble of Regression Trees [11].
- Distance between the face and the cameras estimation. The distance was evaluated by comparing an average anatomical model of a face with the observed data from the calibrated visible camera.
- ROIs calculation based on the landmarks' geometry and signal extraction by taking basic image statistics for each ROI (minimum and maximum, mean and standard deviation of the temperatures of the pixel in the ROI). The assessed ROIs were the tip of the nose, nostrils, glabella and perioral areas.

The classification of the infant's internal state and engagement was built on foundational studies linking the human psychophysiological states and the modulation of nose tip temperature, with whereupon a decrease of temperature dynamic is linked with a sympathetic-like response, associated with distress or negative engagement, whereas its increase is due to a parasympathetic prevalence on the subject autonomic state, related with interest and positive engagement, while a stable dynamic is linked to a neutral engagement [4,12]. Although it was possible to extract signals from different ROIs, only the nose tip has been included in the real-time classification to avoid overloading of the data recording board. Moreover, nose tip has been demonstrated to be the most reliable region for detecting psychophysiological states [5,8].

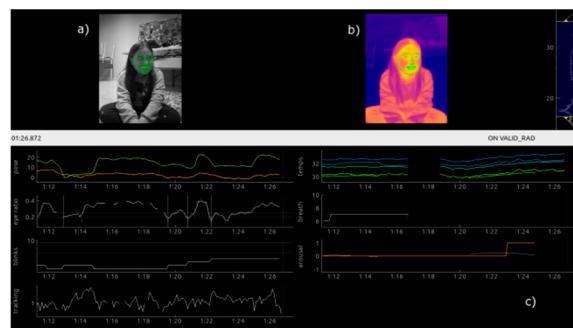


Figure 2. Example of analysis of thermal signals. The coordinates of the landmarks detected in the visible image (a) are reported on the corresponding thermal image (b). The data extracted for the application (c).

3. Results

The system validation was performed by comparing the CPM classification outcome at the end of each event with the corresponding response of the child. Each of the 17 children completed 6 events for a total amount of 102 events. The CPM module recognized a level of interest equivalent to that indicated by the child for 71 events out of 102. The misclassification of the levels of interest was mainly due to artifact movements which leaded to tracking error and noisy signal. Concerning the data extraction, only few samples per video were lost, since not available from visible spectrum camera, whilst an average of 82.75 % of the total amount of the frames was correctly tracked.

4. Discussion

The presented study is aimed to endow a SR with the capability of real time assessment of the interlocutor's state of engagement. By using the described algorithm, with a low resolution and low cost thermal camera, it was possible to understand and classify, with 70% of accuracy, the engagement state of the infant while interacting with an artificial agent. Moreover it was possible to guarantee very high performance on signal processing and a speed of extraction of the signals that went far beyond that of sampling rate of the thermal sensor (about 20 frames per second on an ARM-based single board computer, or about 70 frames per second on a workstation, against a sampling frequency of less than 9 frames per second for the FLIR ONE thermal camera). This study opens up significant prospective on a reliable interaction between the artificial agent and the child by assessing the psychophysiological and emotional state of the child in real time and in a non-invasive fashion, ensuring to maintain an ecologic condition during measurements. A future improvement, whenever applicable, could be the combined use of fIRI with other vital signs acquired by contact devices or more attuned behavioral analysis, to rely on further data and ground truth on the emotional status.

Author Contributions: Conceptualization, C.F., A.M.; methodology, C.F., D.C., A.M.; hardware, C.F., E.S.; formal analysis, C.F., D.C.; writing—original draft preparation, D.C., C.F., E.S.; writing—review and editing, D.C., E.S., C.F.; supervision, A.M.; project administration, A.M.; funding acquisition, A.M.

Funding: This study was funded by MIUR PON Project Sensing Robot—n. F/050326/01-02/X32 FONDO PER LA CRESCITA SOSTENIBILE (F.C.S.) Horizon 2020—PON I&C 2014/2020 (D.M. 1/06/2016).

Acknowledgments: We would like to show our gratitude to the staff and pupils of the primary school “Centro Educativo Il Girasole” in Canzano (TE)—Italy for their assistance during the experimental sessions. Special thanks to Domenico Bianchi, Christian Sciarretta, Maurizio Prezioso (Ud'Anet srl, Italy) e Davide Lisciani (Lisciani Giochi, Italy) for thier support and advice.

Conflicts of Interest: The authors declare no conflict of interest.

References

- 1 Breazeal, C.; Takanishi, A.; Kobayashi, T. Social Robots that Interact with People. In *Springer Handbook of Robotics*; Siciliano, B., Khatib, O., Eds.; Springer: Berlin/Heidelberg, Germany, 2008; pp. 1349–1369.
- 2 Tanaka, F.; Matsuzoe, S. Children teach a care-receiving robot to promote their learning: Field experiments in a classroom for vocabulary learning. *J. Hum.-Robot Interact.* **2012**, *1*, 78–95.
- 3 Leyzberg, D.; Avrunin, E.; Liu, J.; Scassellati, B. Robots that express emotion elicit better human teaching. In Proceedings of the 6th International Conference on Human-Robot Interaction, HRI 2011, Lausanne, Switzerland, 6–9 March 2011; ACM: New York, USA, 2011; pp. 347–354.
- 4 Merla, A. Thermal expression of intersubjectivity offers new possibilities to human-machine and technologically mediated interactions. *Front. Psychol.* **2014**, *5*, 802.
- 5 Ioannou, S.; Ebisch, S.; Aureli, T.; Bafunno, D.; Ioannides, H.A.; Cardone, D.; Manini, B.; Romani, G.L.; Gallese, V.; Merla, A. The autonomic signature of guilt in children: A thermal infrared imaging study. *PLoS ONE* **2013**, *8*, e79440.
- 6 Buddharaju, P.; Dowdall, J.; Tsiamyrtzis, P.; Shastri, D.; Pavlidis, I.; Frank, M.G. Automatic thermal monitoring system (ATHEMOS) for deception detection. In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05), San Diego, CA, USA, 20–25 June 2005; IEEE: New York, NY, USA, 2005.
- 7 Dowdall, J.; Pavlidis, I.T.; Tsiamyrtzis, P. Coalitional tracking. *Comput. Vis. Image Underst.* **2007**, *106*, 205–219.
- 8 Merla, A.; Cardone, D.; Di Carlo, L.; Di Donato, L.; Ragnoni, A.; Visconti, A. Noninvasive system for monitoring driver's physical state. In Proceedings of the 11th AITA Advanced Infrared Technology and Applications L'Aquila, Italy, 7–9 September 2011; Atti della Fondazione Giorgio Ronchi: Florence, Italy, 2011.
- 9 Bradski, G. The opencv library. *Dr Dobb's J. Software Tools* **2000**, *25*, 120–125.
- 10 King, D.E. Dlib-ml: A machine learning toolkit. *J. Mach. Learn. Res.* **2009**, *10*, 1755–1758.

- 11 Kazemi, V.; Sullivan, J. One millisecond face alignment with an ensemble of regression trees. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 24–27 June 2014; IEE: New York, NY, USA, 2014.
- 12 Ioannou, S.; Gallese, V.; Merla, A. Thermal infrared imaging in psychophysiology: Potentialities and limits. *Psychophysiology* **2014**, *51*, 951–963.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).