Detecting Deception in a Human-Robot Interaction Scenario

Roxana Agrigoroaie
Autonomous Systems and Robotics Laboratory,
U2IS, ENSTA-ParisTech
Palaiseau, France
roxana.agrigoroaie@ensta-paristech.fr

Adriana Tapus
Autonomous Systems and Robotics Laboratory,
U2IS, ENSTA-ParisTech
Palaiseau, France
adriana.tapus@ensta-paristech.fr

ABSTRACT

Our work aims at developing a behavioral model enabling customized and natural interaction between a robot and an elderly individual in the context of the EU H2020 ENRICHME Project - Grant Agreement 643691. The main idea of the project is to make the robot able to learn from past experiences and to adapt its behavior to the profile (i.e., personality, cognitive disability level, emotional internal states) and preferences of the person it interacts with. In this paper the robot will try to determine if the person it interacts with is trying to deceive it or not. Some first preliminary experiments and results are presented.

CCS Concepts

•**Computer systems organization** → Robotics;

Keywords

Human-robot interaction; learning; adaptation; deception;

1. INTRODUCTION

In a personalized interaction scenario between a robot and an elderly user, there might be situations in which the robot has to ask if the user executed a certain action or not (e.g., taking medication). In such cases, the robot can either trust what the user is saying or it can use noninvasive methods to determine if the user is trying to deceive it. In this research, we focus on the second option.

Literature review shows that thermal imaging has been successfully used to measure physiological features like pulse rate [12], breathing rate [4], blood flow [10]. As the control of underlying emotions is difficult [2], facial imaging may provide reliable cues to deception. According to [6], temperature variations in some regions of interest (ROI) on the face (i.e., the nose, the forehead, the periorbital regions, and the perinasal region) permit the detection of guilt, anxiety, and stress. Another way of determining the emotions of a person is to use action units (AU)[3]. Emotions can be represented by combinations of AUs [8]. However, the detection of AUs can be a challenge in the context of elderly usage, as age-related wrinkles can lead to false positives.

In this paper, we investigate how likely are people to deceive a robot for a physical prize. Moreover, we are interested in determining if facial temperature variation can be a good indicator of deception.

2. EXPERIMENTAL DESIGN

The experiment that we designed is a pen and pencil-only calculations task (see [9]). The task, entitled "the matrix task" consisted of 25 matrices, each containing 12 decimals on a single sheet of paper. Each participant had 5 minutes to find the two decimals that added up to 10, in as many matrices as they could. Each matrix had one correct answer. For instance, in Table 1 the solution would be 4.81 and 5.19. Participants were instructed to circle each correct answer.

Table 1: Example of matrix

1.69	1.82	2.91
4.67	4.81	3.05
5.82	5.06	4.28
6.36	5.19	4.57

The experiment took place in a laboratory (see Figure 1) with the Meka M1 humanoid robot. This robotic platform has been successfully used in previous experiments in HRI. When entering the room, the robot invited each participant to take a seat at the table found in front of them. Once seated, each participant was told by the robot what the experiment consisted of. Furthermore each participant was instructed to destroy the paper with a paper shredder, set next to the table, before reporting the number of correct answers they got. They were told that nobody will check their results. The paper shredder was modified so as to not completely destroy the answer sheet, which enabled the experimenters to verify the actual number of correct answers, once the participant left the experiment room. For each correct answer they reported, they received one token. After the interaction, the participants could exchange their tokens for physical prizes (e.g., a sandwich, candies, a coffee, etc).

We recorded the RGB and thermal images of the faces of the participants. The RGB image was recorded with an ASUS XTion Live Pro RGB-D camera. The temperature variations across the faces of the participants were recorded with an Optris PI640 thermal camera.

The robot used its left hand to point to the table where the participant should sit, the paper shredder, and the two

HRI '17 Companion March 06–09, 2017, Vienna, Austria

ACM ISBN 978-1-4503-4885-0/17/03.

DOI: http://dx.doi.org/10.1145/3029798.3038376

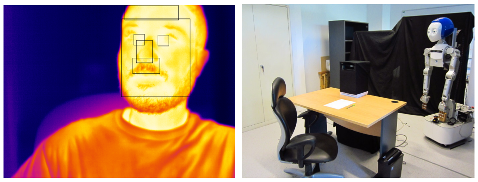

Figure 1: left: the regions of interest shown for one of the participants. right: laboratory setup.

cameras. Meka robot expressed a neutral emotion on its mouth (straight line) with the help of the RGB LED matrix mounted on its head. While each of the participants performed the matrix task, the robot turned its back to him/her so as to not see what was being written on the paper. Once the time was over, the robot turned back towards the participant and asked him/her to check their results and to destroy the paper. When the paper was destroyed the robot asked how many correct answers the participant got. Afterwards the robot told them to take the corresponding number of tokens from the table. Once the experiment was over, each participant had to complete a personality questionnaire based on the Big5 personality traits [5].

3. DATA EXTRACTION

During the interaction, RGB, thermal and audio data was saved using the rosbag utility tool, which is part of the Robot Operating System (ROS)[11]. The purpose of the RGB data was to extract the AUs with a ROS package previously developed in our laboratory. We also developed a ROS package to determine the frequency of the AUs and their combinations, respectively. This enables us to automatically determine if one of the previously mentioned emotions is present during the interaction. The audio data was used to determine the reported answers by the participants. For the thermal data analysis, we defined 6 ROIs on the face: the forehead, the face region, the left, and right periorbital regions, the nose and the perinasal region (see Figure 1). Based on [6], these are the regions where temperature variations could indicate the emotional state of a person. In order to select the ROIs, we apply a face detection algorithm ([7]) on the thermal image. Using [7], we detected 68 facial landmarks and defined our ROIs. We used a correlation tracker (also using [7]) to track the ROIs for the duration of the interaction. From each ROI, we extracted the average temperature. According to [1] the temperatures corresponding to the background or to the presence of glasses had to be removed. We applied the same method for discarding that data.

4. EXPERIMENTAL RESULTS

A total of 51 participants (39 male and 12 female) agreed to take part in this experiment. 26 participants interacted with the robot as described in Section 2, the robot group, while 25 participants were part of the control group. The purpose of the control group was to see what was the average number of matrices a person could solve in an uncontrolled environment.

For the control group the average number of correctly solved matrices was of 9.68 (SD=4.63, min=1, max=21), compared to the average reported answer of 9.88 (SD=4.77, min=2, max=23). For the robot group the average number of correctly solved matrices was of 10 (SD=5.11, min=0, max=20) compared to the average reported answer of 11.38 (SD=5.86, min=1, max=25).

We next compared the answers given by each participant with their result sheets. In the robot group, we found 8 participants that reported a different answer. The difference varied between 1 (5 participants) and 20 (1 participant). In the control group we found 4 participants that reported a different answer. The difference varied between 1 (3 participants) and 2 (1 participant).

We performed two one-way analysis of variance (ANOVA). The first analysis was performed on the number of reported answers based on the personality traits, while the second was performed on the difference between the reported answers and the correct answers based on the personality traits. The results of the first analysis approach statistical significance for the personality trait of agreeableness (F=3.439, p=0.07). However the second analysis did not yield any statistical significance for none of the personality traits.

Preliminary results from the analysis of the thermal data suggest that we can use noninvasive physiological data analysis to determine if a person is trying to deceive a robot.

5. CONCLUSION AND FUTURE WORK

In this paper, we showed how to use thermal camera so as to determine deception in HRI. The first preliminary results are promising. We plan to extract the AUs and analyze in more detail the thermal data for each participant. Once the analysis is completed we intend to run the same experiment with a human experimenter in order to see if there are differences in the number of reported answers and the temperature variations in the ROIs.

6. REFERENCES

[1] Agrigoroaie, R., Ferland, F., Tapus, A.: The enrichme project: Lessons learnt from a first interaction with the elderly. ICSR (2016)

[2] DePaulo, B.M., et al: Cues to deception. Psychological Bulletin (2003)

[3] Ekman, P., Friesen, W.: Facial action coding system. In: Consulting Phychologists Press (1978)

[4] Fei, J., Zhu, Z., Pavlidis, I.: Imaging breathing rate in the co2 absorption band. IEEE EMBS (2005)

[5] Goldberg, L.: An alternative 'description of personality': the big-five factor structure. In: JPSP (1990)

[6] Ioannou, S., Gallese, V., Merla, A.: Thermal infrared imaging in phychophysiology: Potentialities and limits. In: Psychophysiology (2014)

[7] King, D.: Dlib-ml: A machine learning toolkit. In: Journal of Machine Learning Research (2009)

[8] Matsumoto, D., Ekman, P.: Facial expression analysis. In: Scholarpedia (2008)

[9] Mazar, N., Amir, O., Ariely, D.: The dishonesty of honest people: A theory of self-concept maintenance. Journal of Marketing Research (2008)

[10] Pavlidis, I., et al: Thermal imaging for anxiety. Proc. IEEE Workshop CVBVS (2000)

[11] Quigley, M., et al: Ros: an open-source robot operating system. In: ICRA Workshop on Open Source Software (2009)

[12] Sun, N., et al: Imaging the cardiovascular pulse. IEEE CVPR (2005)