A Proposal of Emotion Estimation Method in Social Robots using Facial Expression Recognition Model with Graph-based Techniques

NOPPHAKORN SUBSA-ARD[†1], FELIPE YUDI FULINI[†2],
TIPPORN LAOHAKANGVALVIT[†2], KAORU SUZUKI[†2], and MIDORI SUGAYA[†2]

Abstract: Social robots play a crucial role in human-robot interaction, offering potential benefits in various applications, including nursing care and home assistance. Our previous study proposed "a prototype of multi-modal interaction robot based on emotion estimation method using physiological signals", which presented our implementation on using electroencephalography (EEG) and Heart Rate Variability (HRV) as physiological signals to estimate the emotional states of human. However, wearable sensing devices could be inconvenient to use in the real world for nursing care robots or home-use robots because they may irritate wearing. To tackle this problem, we proposed a new approach for emotion estimation method. Various studies have proposed various methods to estimate human emotions such as recognition of facial expressions, postures, tone of voice, and speech, including the integration of those techniques for further classification. In this study, we focus on the facial expression recognition because facial expression can be captured by a camera instead of any wearable devices, which is easy to use and deploy. Our proposed method employed facial expression recognition with a graph-based technique using an open-source MediaPipe framework to extract face landmarks from photos and used Deep Learning technique to build emotion classification from face mesh data. To validate our proposed method, we used FER2013 dataset from Microsoft as a benchmark dataset. Finally, we implemented our method to the prototype of robot. The results show that our proposed method effectively classify emotions from facial expression captured by web camera, and the robot can interact according to the estimated emotions continuously over time.

Keywords: Emotion estimation, Facial expression, Deep learning, Robot

1. Introduction

In recent years, robots were programmed to handle social interaction with human but had no significant improvement on an emotional understanding. In this research, we focus on machine learning models and methods to recognize the emotions of human based on facial expression recognition to help robots to interact with people efficiently while focusing on the "low computational cost".

According to recent research "A prototype of multi-modal interaction robot based on emotion estimation method using physiological signals" [1]. They proposed to use electroencephalography (EEG) and heart rate variability (HRV) to estimate human emotion using wearable sensors. to measure those physiological data. For the implementation of this method in the ROS-based robot, they sent the raw physiological data from the sensors to the robot, analyzed the data into specific emotions, and used them as criteria for the robot to express its emotions to human.

However, wearable sensors could be inconvenient to use in the real world for nursing care robots or home-use robots because they may irritate wearing [2]. To tackle this problem, we proposed a new approach for emotion estimation, which is facial expression recognition, which enables the emotions to be captured by a camera, which is easy to use and deploy. Our objective is to develop a facial expression recognition model that is lightweight enough to be deployable on our robot prototype using Turtlebot3.

†1 King Mongkut's University of Technology Thonburi
†2 Shibaura Institute of Technology

2. Methodology

2.1 Overall Framework of Our Robot Prototype

Fig.1 illustrates the overall framework of our robot prototype. The module of facial expression recognition by the camera has been added as an improvement from [1]. To achieve this module, we focus on the construction of an emotion recognition model and its application to the robot. To implement the model and apply it to the robot, the pipeline has been designed as shown in Fig. 2. The details are described in the sections 2.2-2.4

Fig. 1 Our pipeline for facial expression-based emotion recognition and application to the robot.

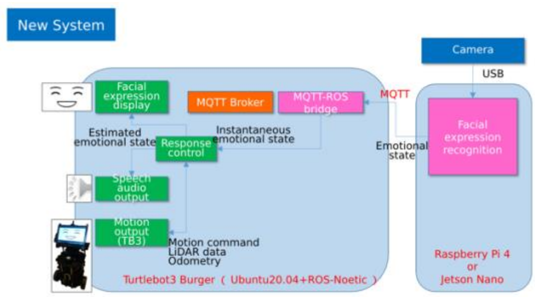

Fig. 2 Our proposed robot framework.

2.2 FER2013 dataset

We used FER2013 (Facial Expression Recognition) benchmark dataset from Microsoft [3] to train, validate, and test our emotion recognition model. The dataset contains 32,298 facial RGB images (48×48 pixels) of different expressions, and the labels divided into 7 types of emotions angry, disgust, fear, happy, sad, surprise, neutral. The dataset predefines the images used for training and validation (N=28,709) and testing (N=3,589).

2.3 Face landmark extraction using MediaPipe

Image-based facial data have some problems especially face angles, for example, faces with sad expression tend to look down while those with happy expression tend to look up. This problem makes the classification model focus on face angle rather than its structure and features. To solve this problem, representing graph-based facial data for modeling as has potential to outperform an image-based one.

Therefore, we propose to extract a graph-based facial surface geometry using an open-source face landmark estimator by MediaPipe [4]. The estimator extracts 478 landmarks in 3D (x, y, and z axis; where x represents horizontal position, y represents vertical position, and z represents distance of the point from camera) from single shot image. This estimator is useful for extracting facial structure as a graph-based data and can be run on low-computational devices such as mobile phones or Raspberry Pi, which is suitable to deploy on robotics.

2.4 Construction of Emotion Classification Model

To construct the emotion classification model, we used the basic neural network classifier, that is the Artificial Neural Network (ANN). We used it as a baseline model for this study with 1434 datapoints (478 landmarks x 3 dimensions), 1024 datapoints in the first hidden layer, 2 of 512 hidden layers, and 7 classes for output layers. Each of layer have ReLU activation and dropout (p=0.2) except the last 7 output layers, optimized with Adam optimizer (lr=3e-4). The model was trained with 1000 epochs. The model architecture is shown in Fig. 3.

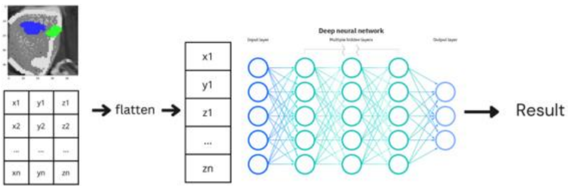

Fig. 3 Our proposed model architecture

2.5 Emotion Mapping to Russell's Circumplex Model

The previous study [1] proposed the robot prototype with a two-dimensional (i.e., arousal and valance) emotion estimation model [5,6] and determined the criteria for robot feedback based on that model. Following the same criteria, we mapped 7 emotions from FER2013 dataset into the same dimensions as follows: happy and surprise to high arousal, high, valence; disgust, anger, and fear to high arousal, low valence; sadness to low arousal, low valence; and neutral to low arousal, low valence.

3. Results and Discussion

Our model yielded only 0.52 accuracy by testing sets, which is undoubtedly not high due to a very simple model architecture. Fig. 5 shows confusion matrix which suggests that the model cannot classify negative emotions (i.e., angry, fear, and sad) from neutral. We need to consider improving the model with more complex architecture to increase the accuracy.

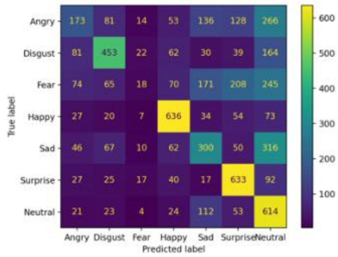

Fig. 5 Confusion Matrix of test set data (1000th epoch)

According to the results, our approaches are working well for positive emotions: happy, surprise, and neutral are outstanding. However, negative emotions are hard to separate from neutral. It may be caused by low quality ground truth label [7]. Our future work will further investigate other datasets for facial expression recognition.

4. Conclusion

Social robots play a crucial role in human-robot interaction, offering potential benefits in various applications. Our previous study proposed an emotion-aware robot. Our research proposes a method to estimate emotions by facial expression recognition model with graph-based technique, that is lightweight to be deployed on the robot. As a result, we successfully deployed the model on the robot. In addition, the model performance itself shows the potential for improvement in various aspects.

Our future work will continue to improve the model that is applicable to social robots. We will also perform a detailed evaluation on the model as well as computational performance on the robot.

Reference

[1] Suzuki, K., Iguchi, T., Nakagawa, Y., & Sugaya, M.: A prototype of multi-modal interaction robot based on emotion estimation method using physiological signals, in Proc. of Asia Pacific Conference on Robot IoT System Development and Platform, vol. 2022, pp. 7–12 (2022).

[2] Ahmad Tarar, A., Mohammad, U., & K. Srivastava, S.: Wearable Skin Sensors and Their Challenges: A Review of Transdermal, Optical, and Mechanical Sensors. *Biosensors*, vol. 10, no. 6, 2020, p. 56 (2020).

[3] I. J. Goodfellow et al.: Challenges in Representation Learning: A report on three machine learning contests", arXiv [stat.ML] (2013).

[4] Kartynnik, Y., Ablavatski, A., Grishchenko, I., & Grundmann, M.: Real-time Facial Surface Geometry from Monocular Video on Mobile GPUs, CoRR, vol. abs/1907.06724 (2019).

[5] Ikeda, Y., Horie, R., & Sugaya, M.: Estimate Emotion with Biological Information for Robot Interaction, KES-2017, Vol.112, pp.1589-1600 (2017).

[6] Yoshida, R. & Sugaya, M.: Emotion Estimation using Biometric Information and its Application, IPSJ SIG Technical Reports, Vol.2019-MBL-90, No.13, pp.1-5 (2019). (in Japanese)

[7] Barsoum, E., Zhang, C., Ferrer, C. C., & Zhang, Z.: Training Deep Networks for Facial Expression Recognition with Crowd-Sourced Label Distribution", in ACM International Conference on Multimodal Interaction (ICMI) (2016).

Acknowledgments This research was a part of International Research Internship Program at the Shibaura Institute of Technology.