Multi-Task Deep Learning based Non-Verbal Communication Method for Cognitive Human-Robot Interaction

KwangEun Ko(kke0217@kitech.re.kr), In Hoon Jang, Gi Hun Yang
HyeunSeog Choi, Bummo Ahn, Dong Wook Lee

Korea Institute of Industrial Technology

Abstract

A non-verbal communication in the process of interaction between human and robot is very important for recognizing information related to implicit intentions, such as facial expressions and gaze, as well as semantic speech recognition, which can transmit explicit command. If the robot can recognize the non-verbal communication information, it can provide a breakthrough of the social robot technology that can create a new service market with a large social and economic potential beyond the current level of providing services as a simple commander (Mavridis, 2015). In this study, we propose a research to recognize facial expressions, eye-gaze, and head gestures important for affective and collaborative interaction between human and service robot. At first, we construct a 2D Convolutional Neural Network (CNN) -based model that recognizes facial expressions, gaze and head direction information based on the data-driven spatial features derived from individual frame image including face regions. And, 3D CNN-based model is used to estimate head movements based on temporal features such as optical flow extracted from consecutive frames (Carreira & Zisserman, 2017). The proposed deep learning-based non-verbal communication information recognition process is completed by weighted fusing the two results. The backbone architecture of the CNNs is based on ResNet-50 (He, Zhang, Ren, & Sun, 2016) and constructed for multi-task learning which can recognize facial expression, eye-gaze, and head direction simultaneously. Also, we tried to improve the accuracy of each recognition task by applying stochastic filter to the recognition result for each time step. The result of the proposed method showed better recognition rate than the result of the deep learning model learned for the individual purpose and the hand-crafted feature based approaches.

Keywords: *non-verbal communication recognition, multi-task learning, facial expression recognition, eye-gaze detection, head-gesture recognition, deep learning*

References

[1] Mavridis, N. (2015). A review of verbal and non-verbal human–robot interactive communication. Robotics and Autonomous Systems, 63, 22–35. https://doi.org/10.1016/J.ROBOT.2014.09.031

[2] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770–778).

[3] Carreira, J., & Zisserman, A. (2017). Quo Vadis, Action Recognition? A New Model and the Kinetics Dataset. Retrieved from http://arxiv.org/abs/1705.07750