Facial Expression Recognition on The Classroom Environments

Wawan Setiawan[1], Yaya Wihardi[2], Enjun Junateti[3], Naufan Rusyda Faikar[4]
{wawans@upi.edu[1], yayawihardi@upi.edu[2], enjun@upi.edu[3]}

Universitas Pendidikan Indonesia, Jl. Dr. Setiabudhi 229, Bandung, Indonesia[1.2.3,4]

Abstract. Facial expression recognition is the process of identifying the expression that is displayed by a person. It can be used to evaluate the mood of students during a class so that can help teachers improve the learning goal achievement. However, the recognition process in real environments such as in classrooms is still a challenging problem due to different expressions and illumination under arbitrary poses. In this paper, we present a convolutional neural network-based method that combining with Gabor filter. The result shows that the proposed method can recognize three categories of student facial expressions that represent a good, bad, and neutral expression.

Keywords: Facial expression, GCNs, Gabor filter, Mood detection.

1 Introduction

Facial expression playing an important role in human social communication, including in classroom teaching. Student facial expressions can express their emotions and indicate their level of interest in the learning process [1]. By recognizing and analyzing them, we can evaluate the students' engagement during a class, so that it can help teachers improve the learning goal achievement.

Studies on facial expressions have been conducted since the 20th century [2]. Recently deep learning-based methods, especially Convolutional Neural Network (CNN), have attracted significant attention in the field of computer vision. Some implementation of CNN in facial expression recognition is demonstrated in [3], [4]. For pose-invariant, Lai et al. [5] proposed a Generative Adversarial Network (GAN), based model. However, the recognition process in real environments such as in the classrooms is still a challenging problem due to different expressions and illumination under arbitrary poses [6].

In this research, we propose a system using Gabor Convolutional Networks (GCNs) model [7] for analyzing students' facial expressions. The students in the classroom get captured by the camera, and then each student's facial region detected by Viola-Jones (V&J) face detector [8]. Finally, the GCNs predicts the classes of students emotion with the highest probability as the resulted one.

MSCEIS 2019, October 12, Bandung, Indonesia

DOI 10.4108/eai.12-10-2019.2296519

2 Related Work

There are two main components that construct the facial expression recognition system: face detector and facial expression classifier.

2.1 Face Detector

To analyze facial expression, the first step to be done is face detection and localization. It segmented face from the rest images so that it becomes easier to classify. Various method can be applied to do this task such as in [8], [9],[10], but based on the survey [11], Viola-Jones (V&J) face detector [8] still considered as an effective one, since it gives a low false-positive rate [11]. In this research, we use the OpenCV version of the Viola-Jones algorithm by adopting in-build class 'Haar feature extraction'.

2.2 Facial Expression Classifier

The study on facial expression classifier is an active area in computer vision. In the conventional system, they used the hand-crafted feature as an input for classifier as shown in [12], [13]. But it only deals with the ideal environment such as good illumination, certain pose, and any ideal condition. For the wild or real environments, it still needs many improvements.

Currently, researchers involve deep learning methods to deal with variation in real environments. [3], [4] implement a CNN-based method to handle this problem. Moreover to cover the lack of data, [5] implement generative adversarial networks for data augmentation so that they can improve the performance of the system. On the other hand, the researcher reported that the Gabor filter can improve the accuracy of the image recognition system by combining it with CNN [7]. Facial expression recognition can be formulated as specific image recognition, so it can be implemented in this problem domain. The flow can be seen on **Figure 1**.

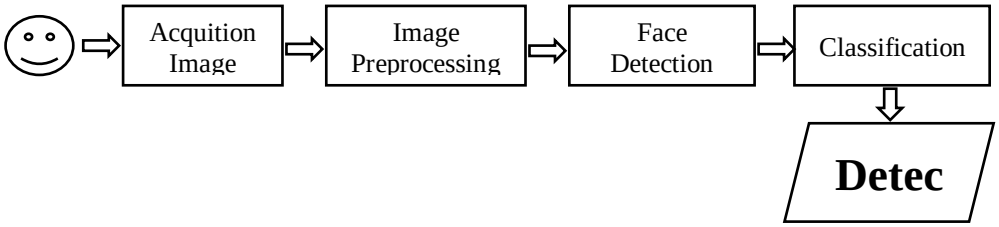

Fig. 1. The flow of Facial Expression Recognition System.

3 Methods

The focus of this work is to build a facial expression system in the real environment, especially the classroom. Firstly we capture the students' faces by using CCTV that installed in front of the class. After that, we preprocess the acquired images to enhance the quality by

doing histogram equalization. The enhanced image becomes an input to the face detector module that implemented by using Viola-Jones (V&J) face detector [8], and the output of this process is the list of detected faces. Finally, each detected face classified by facial expression classifier. These all process are shown in **Figure 1**.

In the classification process, we implement the GCNs model that used in [7], and also implement a standard CNN for comparison. Technically in the implementation, we utilize Tensorflow and Keras to build the model. To train the classifier we use 1890 hand-crafted data gathered from 70 unique people in a real classroom environment. Each person expresses three different expressions that represent a good mood, neutral mood, and bad mood during a class and captured in some different poses.

4 Results and discussion

To validate our method, we do 3-fold cross-validation and use accuracy as a proxy to evaluate the result. Based on the experiments, we obtain 75.71% accuracy of the GCNs model and 67.14% for CNN one as described in Table 1.

Table 1. Comparison The Accuracy of CNN and GCNs.

Methods	Accuracy (%)
CNN	67.14
GCNs	75.71

For qualitative evaluation, we analyze the result on a tested image as shown in **Figure 2**. Based on the result show that it is hard to differentiate the neutral expression and the bad one, due to the high visual similarity between these two classes. To verify our findings, then we compare it with the actual confusion matrix as shown in Tabel 2. There are many misclassified in neutral and bad class, it is contrasted with the good one.

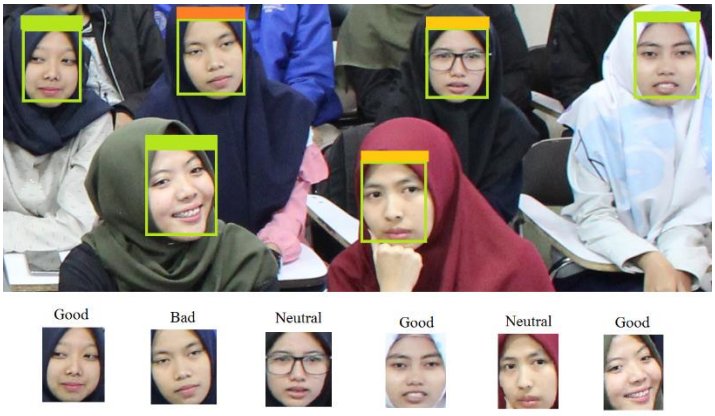

Fig. 2. Detected and Classified Facial Expression Using GCNs.

Table 2. Confusion Matrix for GCNs's Testing.

		Predicted		
		Good	Neutral	Bad
Actual	Good	187	14	9
	Neutral	4	138	68
	Bad	7	51	152

5 Conclusion

We have tried to use GCNs to recognize facial expressions in classroom environments. It can recognize three categories of student facial expressions that represent a good, bad, and neutral expression. The model performance outperforms the original CNN, although it still needs improvement in terms of accuracy.

References

[1] K. S. Sahla and T. Senthil Kumar, "Classroom Teaching Assessment Based on Student Emotions," Springer, Cham, 2016, pp. 475–486.

[2] P. Ekman, "Facial Expression and Emotion," *Am. Psychol.*, vol. 48, no. 4, pp. 384–392, 1993.

[3] H. Ding, S. K. Zhou, and R. Chellappa, "FaceNet2ExpNet: Regularizing a Deep Face Recognition Net for Expression Recognition," in *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)*, 2017, pp. 118–126.

[4] B. Hasani and M. H. Mahoor, "Facial Expression Recognition Using Enhanced Deep 3D Convolutional Neural Networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2017, pp. 2278–2288.

[5] Y.-H. Lai and S.-H. Lai, "Emotion-Preserving Representation Learning via Generative Adversarial Network for Multi-View Facial Expression Recognition," in *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, 2018, pp. 263–270.

[6] F. Zhang, T. Zhang, Q. Mao, L. Duan, and C. Xu, "Facial Expression Recognition in the Wild," in *2018 ACM Multimedia Conference on Multimedia Conference - MM '18*, 2018, pp. 126–135.

[7] S. Luan, C. Chen, B. Zhang, J. Han, and J. Liu, "Gabor Convolutional Networks," *IEEE Trans. Image Process.*, vol. 27, no. 9, pp. 4357–4366, Sep. 2018.

[8] P. Viola and M. J. Jones, "Robust Real-Time Face Detection," *Int. J. Comput. Vis.*, vol. 57, no. 2, pp. 137–154, May 2004.

[9] H. Li, G. Hua, Z. Lin, J. Brandt, and J. Yang, "Probabilistic Elastic Part Model for Unsupervised Face Detector Adaptation," in *2013 IEEE International Conference on Computer Vision*, 2013, pp. 793–800.

[10] V. Jain and E. Learned-Miller, "Online domain adaptation of a pre-trained cascade of classifiers," in *CVPR 2011*, 2011, pp. 577–584.

[11] S. Zafeiriou, C. Zhang, and Z. Zhang, "A survey on face detection in the wild: Past, present, and future," *Comput. Vis. Image Underst.*, vol. 138, pp. 1–24, Sep. 2015.

[12] P. S. Aleksic and A. K. Katsaggelos, "Automatic Facial Expression Recognition Using Facial Animation Parameters and Multistream HMMs," *IEEE Trans. Inf. Forensics Secur.*, vol. 1, no. 1, pp. 3–11, Mar. 2006.

[13] I. Kotsia and I. Pitas, “Facial Expression Recognition in Image Sequences Using Geometric Deformation Features and Support Vector Machines,” *IEEE Trans. Image Process.*, vol. 16, no. 1, pp. 172–187, Jan. 2007.