Facial Expression Recognition System Using Machine Learning

Sanghyuk Kim, Gwon Hwan An, and Suk-Ju Kang
Department of Electronic Engineering
Sogang University
Seoul, Republic of Korea
hitboy91@gmail.com, joviahn@gmail.com, sjkang@sogang.ac.kr

***Abstract*—This paper proposes a novel facial expression recognition system based on image features. There are two main processes in the proposed system, which are face detection and facial expression recognition (FER). The face detection process uses Haar-like features, and the region of interest is reset to reduce the variable of appearance changes. The FER process extracts histogram of oriented gradients (HOG) features from each facial region, and then, support vector machine is performed to classify the final facial expression. In the experimental results, the system exactly recognized the facial expression of a certain person, and the proposed system had the F1 score of 0.8759.**

***Keywords*—machine learning; supervised learning; facial expression recognition**

I. Introduction

In computer vision, facial expression recognition (FER) is a research field that has been studied extensively due to its imperative applications for human-computer interaction, medical treatment, and virtual reality [1]. The FER research can be divided into facial expression detection [2] and facial muscle action detection [3]. In general, the emotional state is defined as 7 emotional states, which are neutral, happy, sad, angry, fearful, surprised and disgusted [4]. FER can be used to analyze the human state such as in the driving situation. Specifically, motion slowdown due to the fatigue is related with traffic accidents [5], and the driver fatigue can be predicted based on the facial expressions of drivers. However, since individual facial expressions can be different, FER requires an individually optimized system.

In this paper, we propose a FER system using machine learning which constructs multi-layer classifiers based on the data set with 7 people. In this case, Haar-like features [6] and histogram of oriented gradients (HOG) features [7] are extracted from each facial region. Then, the support vector machine (SVM) [8] is used to recognize the facial expression of a certain person based on these extracted features.

II. Proposed System

The proposed system is shown in Fig. 1. A facial region is detected by Haar-like features. Then, the facial region of interest (ROI) is reset. HOG features are extracted from the new facial ROI. FER is performed on the extracted HOG features based on SVM. At the face detection and FER processes, if there is no classified region, output will be the same as the input. Each person's data set is composed of the images of angry, happy, and neutral states. The detail processes are explained as follows.

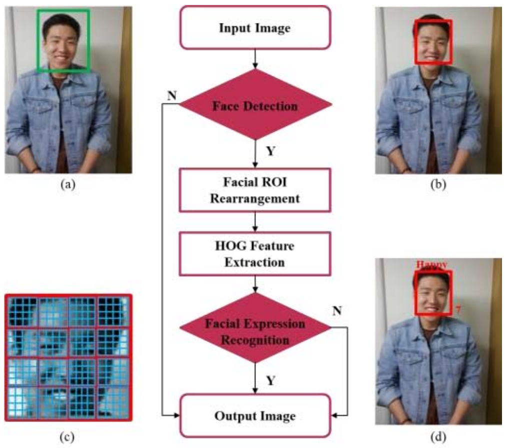

Fig. 1. Flow chart of the proposed system and results of (a) face detection, (b) facial ROI rearrangement, (c) HOG feature extraction, and (d) output image.

A. Face detection

Based on the Haar-like features of an input image, a facial region like the green rectangular in Fig. 1 (a), 64 by 64 pixels, is detected. In the ROI rearrangement, the new facial ROI is determined as follows:

$$(x_{new}, y_{new}, w_{new}, h_{new}) = (7, 20, 50, 38) \quad (1)$$

where x_{new} and y_{new} denote the new pixel starting points of x and y coordinate, and w_{new} and h_{new} denote the new width and height in the ROI. After the ROI rearrangement, the ROI is resized to 32 by 32 pixels to extract HOG features as shown in Fig. 1 (c) only for the face contour such as the red rectangular in Fig. 1 (b). There are two main reasons. First, it is to consider the changeable appearance of each person, for instance, hair style change. Second, the redundant region like a background are minimized in order to improve the classification rate.

TABLE I. ACCURACY AND F_1 SCORE OF FACIAL EXPRESSION RECOGNITION FOR THE CONVENTIONAL(CON.) AND PROPOSED(PRO.) SYSTEMS

	True image input (number)				False image input (number)				Precision		Recall		F_1 score	
	True positive		True negative		False positive		False negative							
System	Con.	Pro.	Con.	Pro.	Con.	Pro.	Con.	Pro.	Con.	Pro.	Con.	Pro.	Con.	Pro.
Angry	76	78	17	15	8	7	123	124	0.8172	0.8387	0.9048	0.9176	0.8588	0.8764
Happy	61	63	17	15	4	2	142	144	0.7821	0.8077	0.9385	0.9692	0.8532	0.8811
Neutral	42	46	11	7	9	7	162	164	0.7925	0.8679	0.8235	0.8679	0.8077	0.8679
Total	179	187	45	37	21	16	427	432	0.7991	0.8348	0.8950	0.9212	0.8443	0.8759

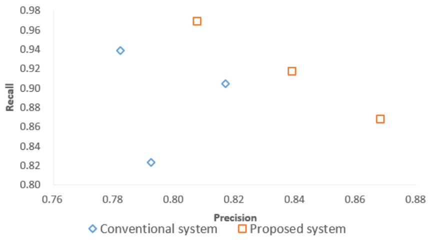

Fig. 2. Graph for the precision-recall relationship in the conventional and proposed system.

B. Facial expression recognition

There are two classifiers which are a person classifier and facial expression classifier. Regardless of the facial expression, in the person classifier, SVM labeled as HOG features of each person is used. The horizontal and vertical gradients of HOG for the facial ROI are calculated as follows:

$$
\begin{aligned}
G_H &= [-1 \quad 0 \quad 1] * B_{ROI} \\
G_V &= [-1 \quad 0 \quad 1]^T * B_{ROI}
\end{aligned}
\tag{2}
$$

where B_{ROI} denotes a ROI block, and G_H and G_V denote the horizontal and vertical gradients based on a 1D-centered discrete derivative mask [7]. The person classifier is constructed by 1:3 ratio of facial image data set to non-facial image data set. Since the existence of faces in an input image is recognized at the person classifier, the facial expression classifier is trained by HOG features of only personalized face images without non-facial images based on SVM.

III. EXPERIMENTAL RESULTS

The experimental results were shown in Table 1. The F_1 score [9] was measured by using a test set. The conventional system combined the person classifier and the facial expression classifier. Therefore, each person's expression was labeled in one layer. The conventional system had the total F_1 score of 0.8443 as shown in Table 1. Even though the facial expression classifier was only for a certain person, at the classifier of the conventional system, facial expression information of both individual and others influenced FER. In order to eliminate the influence of other facial information in FER of a certain person, the proposed system divided person and facial expression classifiers in cascade form. Consequently, two criteria functions independently affected the result of each classification. The proposed system had the F_1 score of 0.8759 as shown in Table 1, which was 0.0316 higher than the F_1 score of conventional system. Fig. 2 represented the data distribution of the precision-recall relationship. In the Fig. 2, the x-axis and the y-axis denoted precision and recall. It showed the same results where both the precision and recall values of the proposed system were higher than those of conventional system.

IV. CONCLUSION

In this paper, we proposed a FER system using machine learning. ROI rearrangement process minimized the environmental change factor and the hierarchical structure of the person, and facial expression classification improved the classification rate. Based on this study, it could be applied to the personalized vehicle interfaces to prevent traffic accidents by combining this proposed system with vehicle.

ACKNOWLEDGMENT

This work is supported by the Korea Agency for Infrastructure Technology Advancement(KAIA) grant funded by the Ministry of Land, Infrastructure and Transport (Grant 17CTAP-C114672-02).

REFERENCES

[1] A. Uçar, Y. Demir and C. Güzeliş, "A new facial expression recognition based on curvelet transform and online sequential extreme learning machine initialized with spherical clustering," Neural Computing and Applications, vol 27, no. 1, pp. 131-142, 2016.

[2] T. Wu, M. S. Bartlett and J. R. Movellan, "Facial Expression Recognition Using Gabor Motion Energy Filters," Computer Vision and Pattern Recognition Workshops, IEEE Computer Society Conference, pp. 42-47, 2010.

[3] J. F. Cohn, Z. Ambadar and P. Ekman, "Observer-based measurement of facial expression with the Facial Action Coding System," The handbook of emotion elicitation and assessment, pp. 203-221, 2007.

[4] M. J. Lyons, J. Budynek and S. Akamatsu, "Automatic Classification of Single Facial Images," IEEE Transcations on Pattern Analysis and Machine Intelligence, vol. 21, no. 12, pp. 1357-1362, Dec. 1999.

[5] J. D. Lee, K. L. Young and M. A. Regon, "Defining driver distraction," in Driver Distraction: Theory, Effects and Mitigation, vol. 13, no. 4, pp 31-40, 2008.

[6] P. Viola and M. Jones, "Rapid Object Detection Using a Boosted Cascade of Simple Features," Computer Vision and Pattern Recognition, IEEE Computer Society Conference, vol. 1, pp. 1-1, 2001.

[7] N. Dalal and T. Bill, "Histograms of Oriented Gradients for Human Detection," Computer Vision and Pattern Recognition, IEEE Computer Society Conference, vol. 1, pp. 885-893, 2005.

[8] C. Corinna and V. Vapnik, "Support-vector Networks," Machine Learning, vol. 20, no. 3, pp. 273-297, 1995.

[9] Y. Yang and X. Liu, "A re-examination of text categorization methods," Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 42-49, 1999.