Facial Emotion Recognition Using 3D Face Reconstruction

Alwin Poulose[1], Chinthala Sreya Reddy[2,3], Jung Hwan Kim[1] and Dong Seog Han[1,*]

[1]School of Electronic and Electrical Engineering, Kyungpook National University, Daegu, Republic of Korea

[2]School of Computer Science and Engineering, Kyungpook National University, Daegu, Republic of Korea

[3]Department of Computer Science, CHRIST University, Bangalore, India

alwinpoulosepalatty@knu.ac.kr[1], sreyareddy2000@gmail.com[2,3], jkim267@knu.ac.kr[1], dshan@knu.ac.kr[1*]

Abstract

In recent days, autonomous driving systems (ADS) effectively utilize facial emotion recognition (FER) results for safe driving. In FER, the system provides the user emotions such as happy, sad, anger, surprise, disgust, fear, or neutral. These emotions provide helpful information for safe driving and reduce the chances of road accidents. The conventional FER approaches use 2D images as their inputs and classify the user emotions. However, the 2D face images in the conventional FER approaches have limited features for model training. In addition, the features from the 2D face images themselves are not sufficient for accurate emotion classification. To reduce the feature extraction issues in the conventional FER approaches, we propose a 3D face image-based FER approach that uses the 3D face reconstruction technique for converting the 2D face images into 3D face images. The deep convolutional neural networks (DCNNs) used in the proposed FER approach efficiently use the 3D face images as inputs and classify the user emotions with minimum errors. The experiment results show that the proposed 3D face image-based FER approach achieves 99% classification accuracy which is better than the conventional 2D face image-based FER approach.

Ⅰ. Introduction

In autonomous driving systems (ADS) [1], the driver's emotions provide crucial information for safe driving [2]. The facial emotion recognition systems (FER) in the ADS recognizes the mental status of the driver and these results reduce the occurrence of road accidents. In FER, the system classifies the driver's emotions into happy, sad, anger, surprise, disgust, fear, or neutral [3]. These classification results give useful information in the ADS for maintaining the vehicle's safety. The FER system captures the driver's emotions in the form of images and classifies the emotions based on the characteristics of the input datasets. In the existing FER approaches, the system uses the 2D faces as input images and the deep learning models used in the FER system classify these emotions based on its training data [4]. However, when a deep learning model [5] uses 2D face images as its inputs, the system needs a huge dataset for model training. As a result, it shows poor performance when a small dataset is used. To address this problem, in this paper, we propose a FER approach that uses 3D face images as the model input and the system shows better classification results with 3D face images than that of 2D images. The classification results from our experiments show that the proposed FER approach reached better classification results than the conventional 2D image-based FER approach.

In this paper, we proposed a 3D face-based FER approach. The proposed system uses the 2D images as inputs and converts them into 3D images using a 3D face construction technique [6][7]. The deep learning model utilizes these 3D face images as its inputs to train the model. The proposed system uses deep convolutional neural networks (DCNNs) as the deep learning model and the 3D face images provide accurate facial features to the DCNNs. To validate our FER approach, we compared the proposed FER approach with the conventional FER approach. The conventional FER system uses 2D face images as its inputs and uses the same DCNNs configuration as the deep learning model. We compared the model accuracy, precision, recall, F1 score and confusion matrices from the proposed and conventional FER approaches in the experiment results and analysis section. The results from our proposed FER approach validate the significance of our approach for emotion classification.

The remainder of this paper is structured as follows. Section II presents our proposed FER system using 3D face reconstruction. Section III discusses the performance of the proposed FER approach with experiment results and analysis. Finally, section IV concludes this paper with future FER research directions.

Ⅱ. Proposed FER System Using 3D Face Reconstruction

The proposed FER approach effectively utilizes the features from 3D face images. The system converts the 2D face images into 3D face images and the deep learning model uses these 3D face images for its training. The proposed FER system consists of a 2D to 3D converter, facial image threshing (FIT) machine [8], DCNNs as the deep learning model which classifies the emotions. Fig. 1 shows the proposed FER system using 3D face reconstruction.

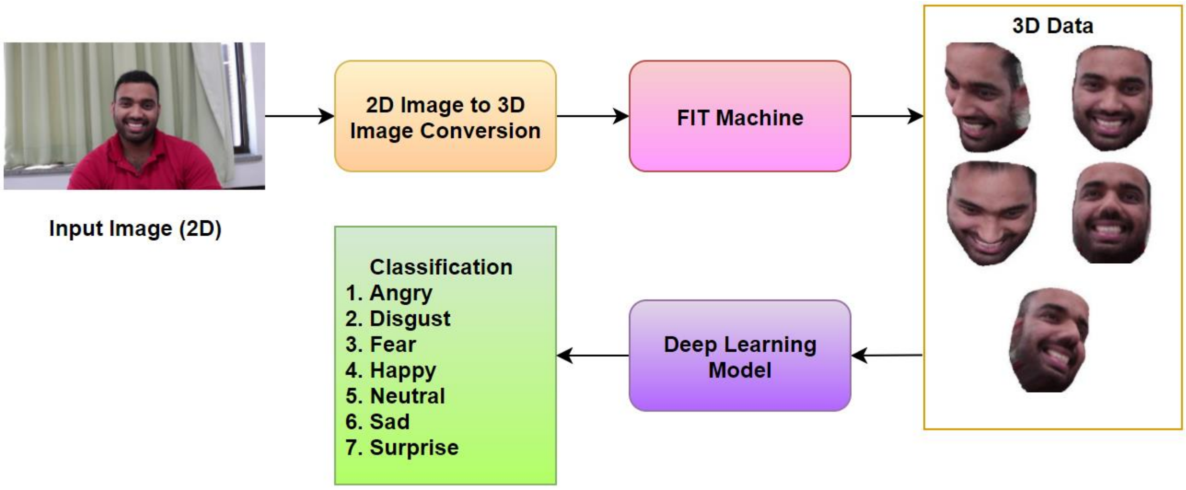


Fig. 1 Proposed FER system using 3D face reconstruction.

The proposed FER system uses the 2D Face images as input and converts them to a 3D dataset using the online 3D face reconstruction tool. The system then pre-processes the 3D face images by the FIT machine and removes the unnecessary features from the dataset. The output of the FIT machine is a resized 3D face image with 48 × 48 resolution. Next, the system converts the color 3D face images into grayscale. Finally, the system splits the grayscale 3D face image dataset into training, testing and validation in the ratio of 60:20:20 and trains the model using the training and validation data. The trained DCNNs model classifies emotions into one of the seven categories: happy, sad, anger, surprise, disgust, fear, and neutral.

Ⅲ. Experiment and Result Analysis

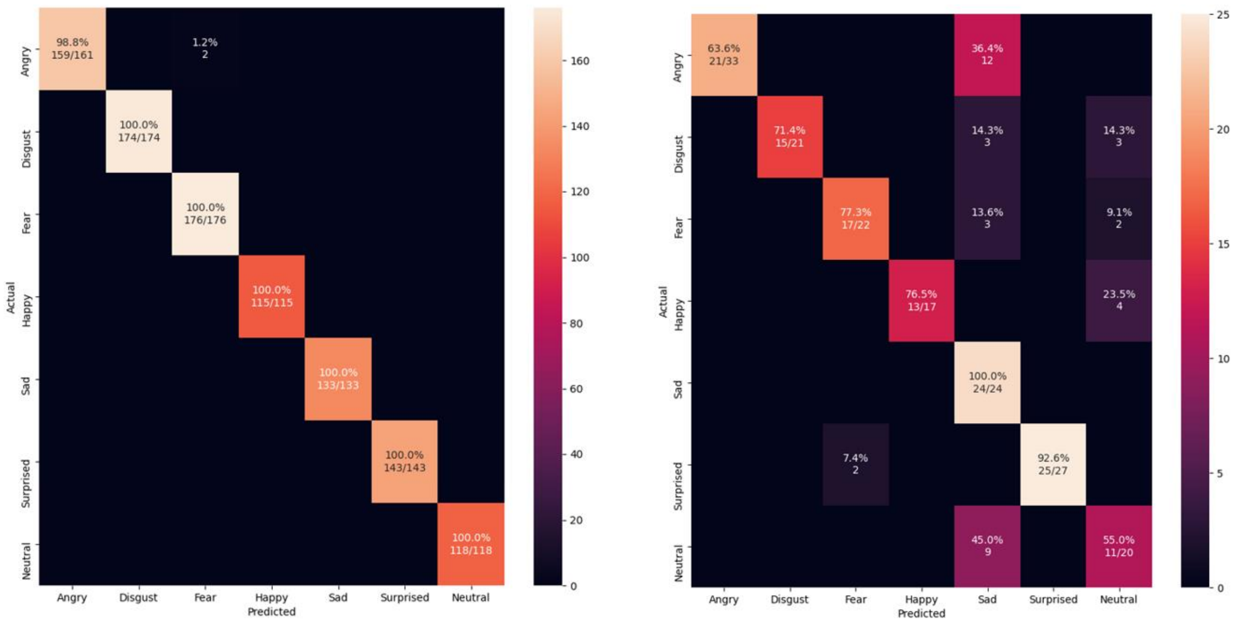
To validate our proposed FER approach using 3D face reconstruction, we did our experiment with a user. First, we simulated the FER environment and collected the data with seven emotions. The dataset consists of happiness, sadness, angry, surprise, disgust, fear, and neutrality. Next, we converted this dataset into 3D face images and used them for the DCNNs training. Fig. 2 shows the performance of the proposed and conventional FER approaches.

From Fig. 2, the proposed 3D face-based FER approach shows better classification results than the conventional 2D face-based FER approach. The proposed 3D face image-based approach improves the recognition rate and the DCNNs model has a 23.16% accuracy improvement compared to the conventional approach. The results from the confusion matrices validate our proposed FER approach and this approach is an effective method for FER applications. The performance of the proposed and conventional approaches is summarized in Table 1.

Table 1: Performance comparison of proposed and conventional FER approaches.

FER Approach	Accuracy (%)	Precision	Recall	F1 Score
Proposed FER	99.8%	99.80	99.80	99.80
Convnetional FER	76.82%	85.35	76.82	78.33

From Table 1, the proposed FER approach achieved a better accuracy, precision, recall and F1 score than the conventional FER approach. These results demonstrate the impact of the 3D face image-based FER approach and our method is a suitable FER approach in ADS. The 3D face images provide better facial features for model training as compared to 2D face images. The DCNNs model trained with a 3D face dataset has more precise classification results than the conventional FER approaches.



(a) (b)

Fig. 2 Performance of the FER approaches. (a) Proposed FER approach. (b) Conventional FER approach.

IV. Conclusion

In this paper, we proposed a FER system using 3D face reconstruction. The proposed FER system takes the 2D face images as the input and converts them into 3D face images. The DCNNs used in the proposed FER utilizes the features of 3D images and provide real user emotions. The experiment results show that the proposed 3D face image-based FER approach achieves more accurate emotion classification results than the conventional 2D face image-based FER approach. In future works, we can extract the face landmarks from the 3D images and improve the performance of the proposed FER approach.

Acknowledgment

This research was supported by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2021-2020-0-01808) supervised by the IITP (Institute of Information & Communications Technology Planning & Evaluation).

References

[1] A. Poulose and D. S. Han, " Hybrid Deep Learning Model Based Indoor Positioning Using Wi-Fi RSSI Heat Maps for Autonomous Application," *Electronics*, vol. 10, p. 2, 2021.

[2] J. H. Kim, A. Poulose, and D. S. Han, " Facial Image Threshing Machine for Collecting Facial Emotion Recognition Dataset," in *Proceedings of the Symposium of the Korean Institute of communications and Information Sciences (KICS) Fall Conference*, Seoul, Korea, 13 November 2020; pp. 67- 68.

[3] J. Rößler, J. Sun, and P. Gloor, " Reducing Videoconferencing Fatigue through Facial Emotion Recognition," *Future Internet*, vol. 13, p. 126, 2021.

[4] Y. Fu, J. Ma, and X. Guo, " A Foreground-Aware Framework for Local Face Attribute Transfer," *Entropy*, vol. 23, p. 615, 2021.

[5] A. Hafner, P. Peer, Ž . Emerš ič , and M. Vitek, " Deep Iris Feature Extraction," in *International Conference on Artificial Intelligence in Information and Communication (ICAIIC),* pp. 258-262, 2021.

[6] A. S. Jackson, A. Bulat, V. Argyriou, and G. Tzimiropoulos, "Large pose 3D face reconstruction from a single image via direct volumetric CNN regression," in *Proceedings of the IEEE international conference on computer vision*, pp. 1031-1039, 2017.

[7] A. S. Jackson, A. Bulat, V. Argyriou, and G. Tzimiropoulos, " Large pose 3D face reconstruction from a single image via direct volumetric CNN regression," in *Proceedings of the IEEE international conference on computer vision,* pp. 1031-1039, 2017.

[8] J. H. Kim, A. Poulose, and D. S. Han, " The extensive usage of the facial image threshing machine for facial emotion recognition performance," Sensors, vol. 21, p. 2026, 2021.