KS-FER: Kernel Sharpening Based Facial Emotion Recognition

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

This paper proposes a kernel sharpening-based facial emotion recognition (KS-FER) system using the deep learning Xception architecture. The conventional FER systems use raw facial images from a camera for emotion recognition. However, the raw images contain background noise which reduces the performance of the FER system. The deep learning models in the FER systems have challenges extracting accurate facial features from the raw input images if the background noise ratio is large. To address the background noise in the facial images, in this paper, we propose a kernel sharpening technique for facial images, and this approach enhances the facial features of the input images. The kernel sharpening accentuates the edges of facial images and adds contrast to them. The kernel sharpening-based facial images provide more accurate input features to the deep learning model than the raw facial input images by removing the background noise from the facial images. Our KS-FER system's experiment and result analysis show that the proposed approach has a 16.54% improvement in accuracy than the conventional FER approach. The proposed KS-FER approach also reduces the loss of the Xception model from 0.58 to 0.027 for emotion recognition.

Ⅰ. Introduction

The facial emotion recognition (FER) system classifies the emotions and assists the driver monitoring systems (DMS) for safe driving [1]. In FER, the system uses facial images as input and identifies the user's emotion based on his/her facial expressions [2]. The results from the FER system are helpful for safe driving as it can reduce road accidents in autonomous driving systems (ADS) [3]. The conventional FER systems use facial images from the cameras and it classifies them into seven user emotions: happiness, sadness, anger, surprise, disgust, fear, and neutrality. When the conventional FER systems use the raw facial images from the camera, the input images are not free from the background noise. The deep learning model in the FER system fails to classify the user emotions accurately due to the lack of proper input features. The background noise in the facial images reduces the classification accuracy, and the classification results from the deep learning model have a higher degree of errors. To address the background noise in the facial images, in this paper, we propose a kernel sharpening-based facial emotion recognition (KS-FER) system for emotion recognition. The KS-FER system reduces the background noise in the facial images and improves the features that the deep learning model can effectively use. The results from the KS-FER system show that this approach is an effective method for emotion recognition.

In this paper, we propose a KS-FER system that can classify seven user emotions with lesser errors. In KS-FER, the kernel sharpening highlights the edges of the facial images and makes the transitioning of facial features more significant to the deep learning model. The kernel sharpening emphasizes the differences in adjacent pixel values that make the facial images more vivid. The final images after kernel sharpening are enhanced facial images with lesser background noise. The deep learning model uses these images as its input to classifies the user emotion. Our FER system uses the Xception architecture as the deep learning model, and the kernel sharpened images provide accurate facial features to this model. The classification results from our KS-FER show that the kernel sharpening-based FER approach reduces the classification error and the Xception model effectively uses the kernel sharpened facial images.

The rest of the paper is organized as follows. Our proposed FER system using kernel sharpening is presented in Section II with a detailed explanation of the KS-FER system. Section III discusses our FER experiments and results with a comparative analysis of proposed KS-FER and conventional FER approaches. Section IV concludes this paper with future research directions.

Ⅱ. Proposed FER System Using Kernel Sharpening

The proposed KS-FER system effectively utilizes foreground extraction [4], kernel filtering, and the facial image threshing machine (FIT) [5] for image preprocessing. Fig. 1 shows the proposed KS-FER system.

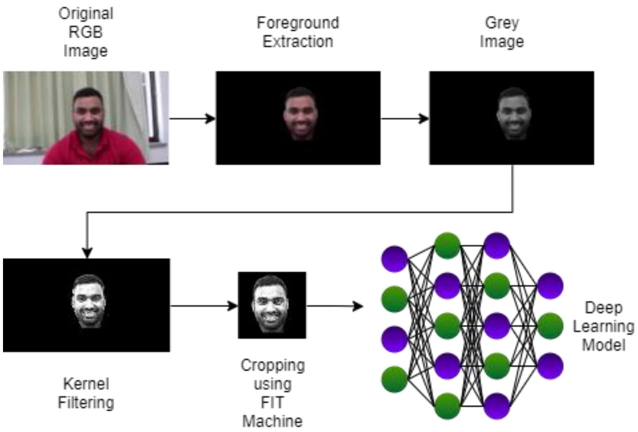

Fig. 1 Proposed KS-FER system.

In the proposed KS-FER system, we start our FER data collection from different users using a camera. The FER dataset contains seven facial expressions which are processed by a foreground extraction technique. The foreground extraction technique removes the background of the images from the raw images and retains the face of the person. After foreground extraction, we convert the images into greyscale which reduces the computational complexity of the system. A kernel sharpening technique is applied to the greyscale images, which enhances the edges of the facial images with significant features. Fig. 2 shows the raw images along with their kernel sharpened images of the different user emotions.

As shown in Fig. 2, the kernel sharpening process highlights the emotions of the user and improves the features for better classification. A FIT machine is used after the kernel sharpening process which resizes and crops the facial images. The output from the FIT machine is a suitable image to feed the deep learning model, and the model can efficiently train with these images. An Xception architecture is used as the deep learning model, which classifies the seven user emotions. Fig. 3 shows the Xception architecture used in our KS-FER system.

As shown in Fig. 3, the Xception model uses an image size of 48×48 as its input with Adam as its optimizer. The model uses 0.02 as its learning rate with a single channel and a batch size of 128. The model is set to train for 100 epochs with the early stopping method. The proposed KS-FER and conventional FER approaches use the same model configuration except for the size of input images. The conventional FER approach uses 88×50 as the input image size as we maintain the aspect ratio of the original image.

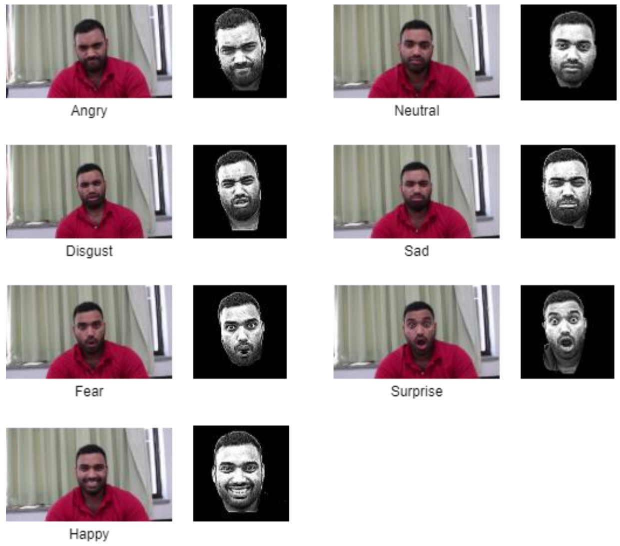

Fig. 2 Kernel sharpening on user emotions.

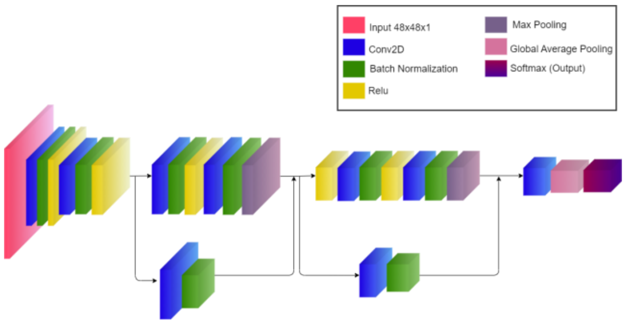

Fig. 3 The Xception model used in the KS-FER system.

Ⅲ. Experiment and Result Analysis

To validate our KS-FER system for emotion recognition, we collected images of facial expressions made by 9 people for the following emotions: happiness, sadness, angry, surprise, disgust, fear, and neutrality. Our FER dataset consists of 9,000 facial images, and we used 60% of the facial images for model training, 20% for validation, and 20% for testing the model. The Xception model used in our KS-FER system uses this dataset, and the confusion matrix results from proposed and conventional approaches are shown in Fig. 4.

As shown in Fig. 4, the conventional FER system uses the raw images without foreground extraction or kernel sharpening and has a higher classification error. The conventional approach achieved an accuracy of 82.59% with a model loss of 0.583. The KS-FER approach achieved 99.1% model accuracy with a model loss of 0.027. These results indicate that the proposed approach has an improvement of 16.54% in accuracy when compared to the conventional FER approach. The performance of our KS-FER and conventional FER approach is summarized in Table 1.

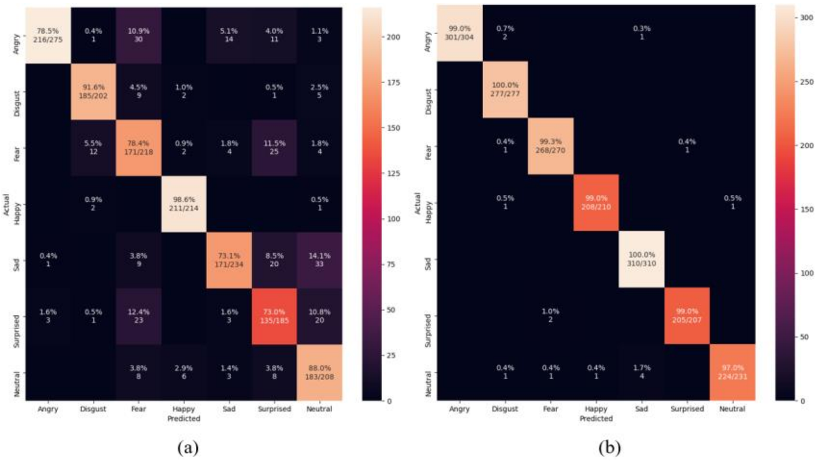

Fig. 4 Performance of the FER approaches. (a) Conventional FER approach. (b) Proposed KS-FER approach.

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

FER Approach	Accuracy (%)	Precision	Recall	F1 Score
Proposed KS-FER	99.13%	99.12	99.11	99.11
Conventional FER	82.59%	84.13	82.81	83.02

From Table 1, we can see that the KS-FER system outperforms than conventional FER approach in terms of accuracy, precision, recall, and F1 score. The kernel sharpening approach improves the features of the facial images, and the deep learning model can easily interpret the expressions in the facial images. The foreground extraction technique reduces the background noise. The kernel sharpening and foreground extraction together enhance the facial features in the images for accurate classification. From the experiment and result analysis, it is clear that the KS-FER approach has a significant improvement in facial emotion recognition systems.

IV. Conclusion

In this paper, we proposed a KS- FER system for emotion recognition. The proposed approach effectively utilizes the foreground extraction and kernel sharpening of the facial images and improves the input features learnt to the deep learning model. The system uses the Xception architecture to classify seven user emotions. The experiments and results show that the KS-FER system improves the classification accuracy and reduces the model loss for emotion recognition. We can add Gaussian blur and unsharp mask image processing techniques to the facial images for better FER performance in future work.

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) and the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2021R1A6A1A03043144).

References

[1] D. Y. Choi and B. C. Song, “Facial Micro-Expression Recognition Using Two-Dimensional Landmark Feature Maps," IEEE Access, vol. 8, pp. 121549-121563, 2020.

[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. H. Kim, R. Mutegeki, A. Poulose, and D. S. Han, "A Study of a Data Standardization and Cleaning Technique for a Facial Emotion Recognition System," In proceedings of the Symposium of the Korean Institute of communication and Information Science (KICS) Summer Conference, pp. 1193-1195, 2020.

[4] C. Rother, V. Kolmogorov, and A. Blake, “GrabCut interactive foreground extraction using iterated graph cuts,” ACM transactions on graphics (TOG) vol. 23, pp. 309-314, 2004.

[5] 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.