Computers, Materials & Continua

Tech Science Press

DOI: 10.32604/cmc.2023.039460
Article

Deep Facial Emotion Recognition Using Local Features Based on Facial Landmarks for Security System

Youngeun An, Jimin Lee, EunSang Bak* and Sungbum Pan*

IT Research Institute, Chosun University, Gwang-Ju, 61452, Korea
*Corresponding Authors: EunSang Bak. Email: bakeunsang@chosun.ac.kr; Sungbum Pan. Email: sbpan@chosun.ac.kr
Received: 31 January 2023; Accepted: 23 May 2023; Published: 30 August 2023

Abstract: Emotion recognition based on facial expressions is one of the most critical elements of human-machine interfaces. Most conventional methods for emotion recognition using facial expressions use the entire facial image to extract features and then recognize specific emotions through a pre-trained model. In contrast, this paper proposes a novel feature vector extraction method using the Euclidean distance between the landmarks changing their positions according to facial expressions, especially around the eyes, eyebrows, nose, and mouth. Then, we apply a new classifier using an ensemble network to increase emotion recognition accuracy. The emotion recognition performance was compared with the conventional algorithms using public databases. The results indicated that the proposed method achieved higher accuracy than the traditional based on facial expressions for emotion recognition. In particular, our experiments with the FER2013 database show that our proposed method is robust to lighting conditions and backgrounds, with an average of 25% higher performance than previous studies. Consequently, the proposed method is expected to recognize facial expressions, especially fear and anger, to help prevent severe accidents by detecting security-related or dangerous actions in advance.

Keywords: Facial emotion recognition; landmark-based feature extraction; ensemble network; robustness to the changes in illumination and background; dangerous situation detection; accident prevention

1 Introduction

Facial expressions reflect various information, including the states of mind, the nature of social interaction, physiological signals, and human emotion. Recently, facial expression-based emotion recognition (hereafter referred to as facial emotion recognition) techniques have emerged as a significant area of interest in computer vision, computer graphics, and human-computer interaction (HCI). In particular, emotion recognition technology using vision-based facial motion tracking and expressions attracts significant attention as an effective modality for HCI [1]. However, conventional facial recognition is a technology that detects people in interest and does not determine whether a person intends to commit illegal, harmful, or dangerous actions.

This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Al-Modwahi et al. [2] developed a real-time facial expression recognition system that can immediately recognize a person's facial expression and notify a security guard before a prohibited action is executed. Sajjad et al. [3] developed a system that analyzes facial expressions to pre-recognize activities such as robbery or fights between people for the intelligent security of law enforcement services. Kim et al. [4] built a system that predicts the patient's movement during treatments by recognizing the patient's emotions. It helps in detecting risks in advance and safe treatment of patients. Kim et al. [5] also identified the patient's condition by analyzing real-time patient data before the patient fell into shock and tried to predict shock and save the patient's life.

The research on facial emotion recognition has become increasingly important since it collaborates with face detection, face tracking, and face recognition. Furthermore, due to technological advancements, fast processing at a low cost also contributes to accelerating the growth of this technology. Analyzing the changes in facial expressions requires an optimal facial information extraction method that can differentiate the changes. However, extracting facial expression information could be difficult, depending on the variation of the illumination and background [1]. If many elements of facial expressions were considered to recognize emotions, the computation would be excessive and time-consuming. Interestingly, people usually infer a specific emotion simply by glancing at another person's face [6,7]. Such experience shows that a small number of particular elements of facial expressions play a crucial role in recognizing emotions.

The eyebrows, eyes, nose, and mouth are landmarks of the face that prominently reveal its characteristics of the face. These landmarks are where the movements of the facial muscles are well-reflected. Rhodes [8] showed that the features of the eyebrows, eyes, nose, and mouth (and the spatial relationships among these features) are more critical for recognizing emotions than other features of the face, such as hair, wrinkles, and moles through several studies on facial emotion recognition. Similarly, the eyebrows, eyes, nose, and mouth are also crucial in facial expression recognition. Pilowsky et al. [9] reported that the distance between the features of a facial expression is essential for recognizing facial expressions. In these studies, two terms, facial expression recognition and facial emotion recognition, are used interchangeably.

Therefore, in this paper, we propose a novel emotion recognition method taking advantage of the features from a partial face to overcome the difficulty of extracting valuable features from the whole area of the face due to their sensitivity to the variation of illumination and background. We used various public databases to evaluate the proposed method's performance compared to the conventional methods.

The remainder of this paper is organized as follows. Section 2 describes the public databases and the conventional facial emotion recognition methods, and Section 3 describes the proposed method. In Section 4, we compare the result of the proposed method with the conventional facial emotion recognition method. Finally, in Section 5, we conclude the paper.

2 Previous Studies on Facial Emotion Recognition

2.1 Public Databases

Among the various public databases, we used CK+ (Extended Cohn-Kanade), JAFFE (Japanese Female Facial Expression), and FER2013 (Facial Expression Recognition 2013), which are the most common databases used in the facial emotion recognition field, in this study. The CK+ and JAFFE databases consist of facial expression data obtained in limited environments with rare changes

in the illumination and background. On the contrary, the FER2013 database consists of facial-expression data with various illumination conditions and backgrounds collected using the Google image search API.

The CK+ database is composed of data from 100 people between the ages of 18 and 30. The ratio of men to women is 35:65, with 15% of the subjects being African-American and 3% Asian or Latino.

Each video consists of 30 frames with a resolution of 640 × 480 or 640 × 490 pixels and is marked with one of seven facial expressions (or emotion): anger, disgust, contempt, fear, happiness, sadness, and surprise. The training and test set consists of 1,390 and 245 examples, respectively. Table 1 presents the classified data in total for each emotion, and Fig. 1 shows sample images.

Table 1: CK+ database configuration

Emotion	Number of data
Anger	225
Contempt	90
Disgust	295
Fear	125
Happiness	345
Sadness	140
Surprise	415
Total	1,635

Figure 1: Examples of CK+ database samples

The JAFFE database contains of 213 images of 10 Japanese women with various facial expressions. Each person was asked to show seven facial expressions (six basic expressions and one neutral expression). Then, 60 annotators examined each facial expression image and selected one facial

expression among the seven to provide annotations. The annotation for which a majority voted became the representative facial expression. The resolution of each image is 256 × 256 pixels. The training and test set consists of 181 and 32 examples, respectively. Table 2 presents the number of data for each emotion, and Fig. 2 shows sample images.

Table 2: JAFFE database configuration

Emotion	Number of data
Anger	30
Disgust	29
Fear	32
Happy	31
Sad	31
Surprise	30
Neutral	30
Total	213

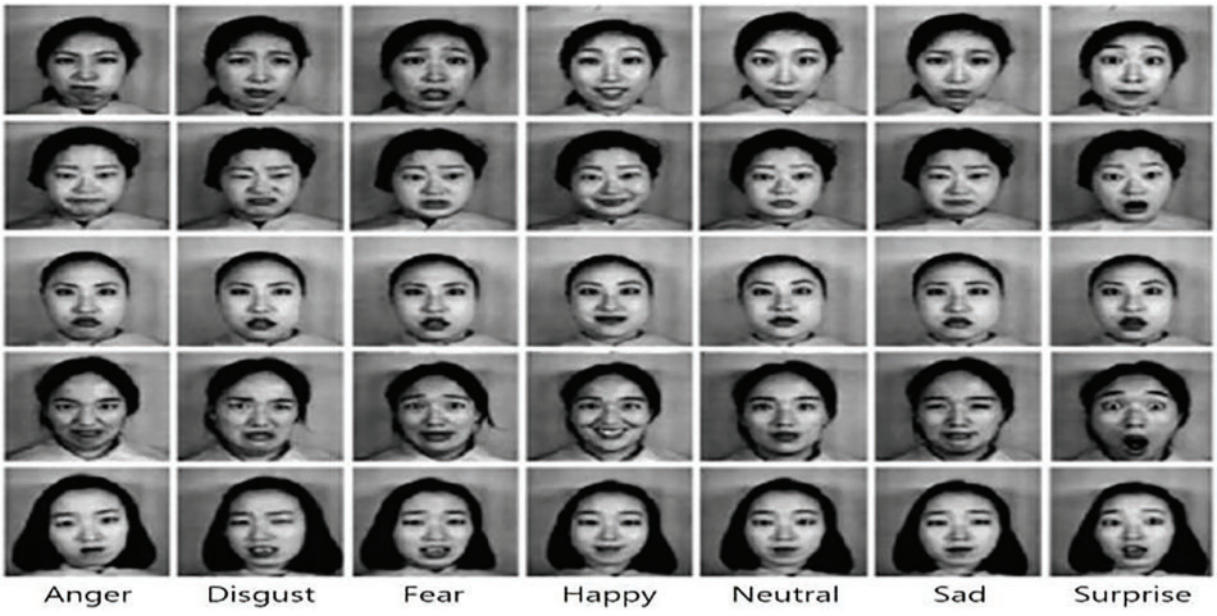


Figure 2: Examples of JAFFE database samples

Lastly, the FER2013 database contains facial expression data with various illumination conditions and backgrounds, and the emotion recognized in each facial expression belongs to one of seven categories (Angry, Disgust, Fear, Happy, Sad, Surprise, and Neutral). Each image has a resolution of 48 × 48 pixels. The training and test set consists of 28,709 and 3,589 examples, respectively. Table 3 presents the number of data for each emotion, and Fig. 3 shows the sample images.

Table 3: FER2013 database configuration

Emotion	Number of data
Angry	4,486
Disgust	491
Fear	4,625
Happy	8,094
Sad	5,424
Surprise	3,587
Neutral	5,591
Total	32,298

Figure 3: Examples of FER2013 database samples

2.2 Conventional Emotion Recognition Methods Using Facial Expressions

The feature extraction methods of facial emotion recognition can be typically divided into two categories: statistical analysis-based and landmark-based feature extraction methods. Extracting features via statistical analysis uses pixel information extracted from the whole face or a partial region of the face and typically uses eigenvectors to extract features. In contrast, the landmark-based feature

extraction method uses the position information of the landmarks, such as the eyes, nose, and mouth. Thus, various algorithms exist depending on the techniques of composing landmarks. Subsequently, different machine-learning or neural network methods are required to recognize emotions using the extracted features [1].

Mollahosseini et al. [10] proposed a novel deep neural network architecture with two convolutional layers using GoogLeNet and AlexNet. Liu et al. [11] used an adaptive three-dimensional (3D) convolutional neural network (CNN) that contained deformable actions to detect a particular facial action part under the structured spatial constraints and simultaneously obtain a discriminatory expression part. Liand et al. [12] developed a learning network called Deep Bi-Manifold CNN (DBM-CNN) by using multi-label cross-entropy losses and adding a new deep layer called the bi-manifold loss layer. In this layer, the adjacent neighbor label vectors are forcibly grouped. This task is repeated to make them more densely aligned. Tang [13] adopted CNNs similar to AlexNet and employed a linear 1-vs-all top layer and linear support vector machines (SVMs) instead of the softmax function. Minaee et al. [14] proposed a technique that detects important regions of the facial expression using a facial expression recognition network to exploit the feature information on a partial area of the face, similar to the method proposed herein. Khaireddin et al. [15] proposed an optimization method based on a VGGNet architecture, which fine-tuned the hyperparameters. Abidin et al. [16] proposed a neural network-based facial expression recognition using Fisherface features. Happy et al. [17] extracted landmarks from the face, similar to the proposed method, and chose the prominent features as patches to use as feature information. In contrast to the proposed method, this method uses the middle of the forehead, cheeks, and lip corners as landmark-based facial feature information. Jain et al. [18] proposed a method for classifying six face graph classes based on a single deep convolutional neural network (DNN) containing convolutional layers and deep residual blocks. Shima et al. [19] proposed a system for extracting facial features based on a deep neural network and an SVM. Jain et al. [20] proposed a hybrid convolution-recurrent neural network method, a network architecture consisting of convolution layers followed by a recurrent neural network (RNN). Such a combined model extracts the relations within facial images and their temporal dependencies by using the recurrent network.

Although many researchers have attempted to recognize emotions using neural networks combined with facial feature extraction methods, the inherent drawbacks of being susceptible to illumination conditions and backgrounds have yet to be fully overcome. Still, relevant research is underway to address this issue in various ways. In the paper, we employed a landmark-based method to overcome the problem. We propose an algorithm that extracts landmark information from the face, generating efficient features, and then uses them to detect changes in facial expressions for emotion recognition. The performance of the proposed algorithm was compared with that of the conventional algorithms, particularly for the FER2013 database, which contains many images of different illumination conditions and backgrounds.

3 Emotion Recognition Using Feature Information of Partial Face Regions

This paper proposes an emotion recognition method using partial information about the face to improve the performance of facial emotion recognition. Fig. 4 shows a block diagram of the proposed system. As shown in Fig. 5, the Haar cascade-based method [21–23], which Viola and Jones proposed, is used as a normalization process to crop only the facial region, as in Fig. 5b, from an original face image. Then we extract feature vectors from the cut area. During the process, the background is removed from the original image, which becomes a face image without the background. Now, the adverse effects of the background are removed from the face image. This procedure enables the

proposed method to be robust against background variations. Next, we extract the landmarks [24] from the facial region as red-filled circles in Fig. 6b. Among the red-filled circles, we filter out a few of the most relevant landmarks to facial expressions connected with yellow lines in Fig. 6b. They are positioned around the eye, eyebrows, nose, and mouth. Finally, we can to construct the feature vectors by measuring the distance between the yellow-lined circles. As in Fig. 6a, the d1-10 from the eye and eyebrows, d11-12 from the nose, and d13-18 from the mouth are the distances, each resulting in a feature vector. Consequently, we are given feature vectors for local characteristics from the landmarks of three different regions as well as the feature vectors from the whole face area for global characteristics. Those four types of feature vectors are employed to train the respective CNN. Subsequently, the soft voting method decides on the recognized facial emotion.

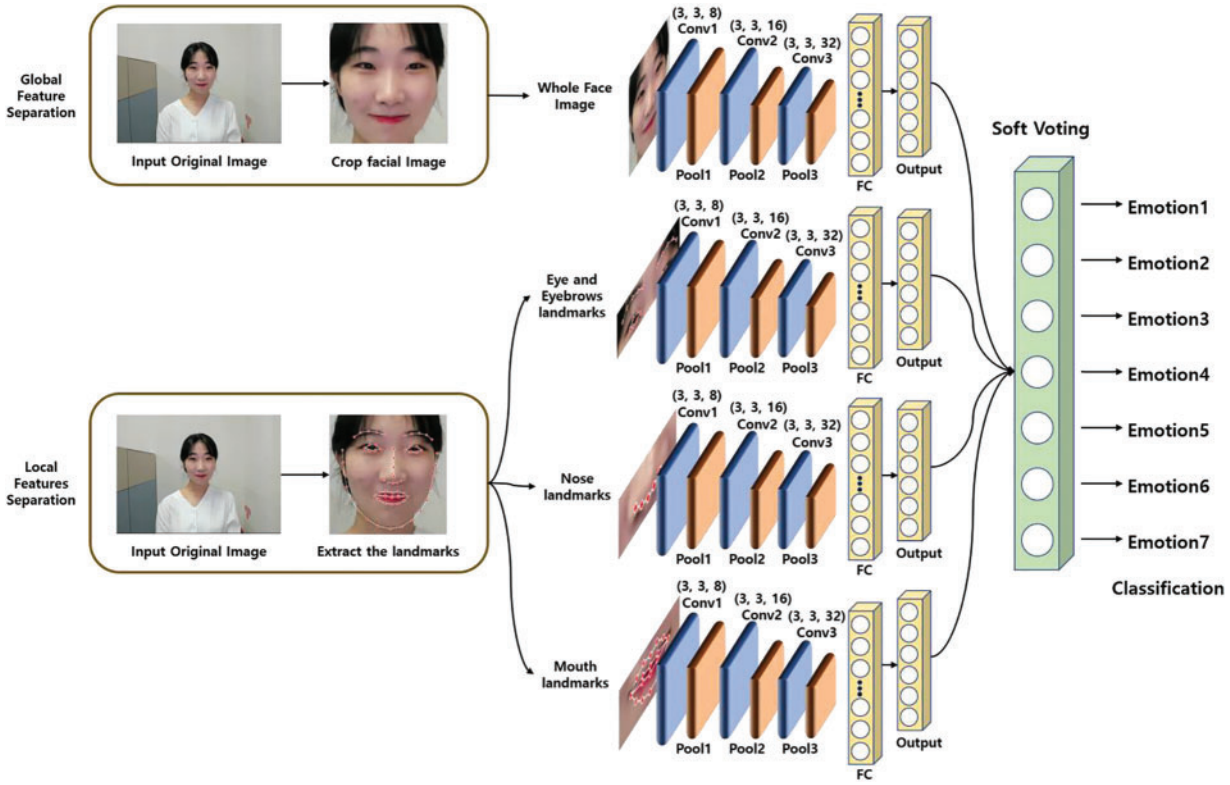
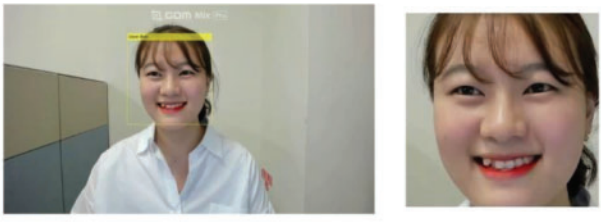


Figure 4: Block diagram of the proposed system



(a) Original image (b) Face image

Figure 5: As a result of the normalization process to crop only the facial region from an original image

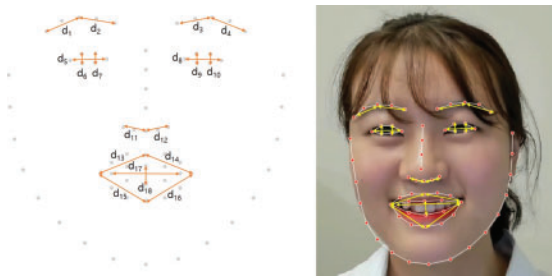


Figure 6: Examples of extracting the landmark distance information

As shown in Fig. 6, the landmarks expressing feature information in the proposed algorithm are related to the eyes, eyebrows, nose, and mouth among the 68 extracted landmarks as the red-filled circles in Fig. 6b. We chose them according to the facial expression-based features [25] presented in Table 4, which were suggested by Paul Eckman, who demonstrated the universality of seven primary facial expressions of emotions found in social and psychological research for several decades (since the 1970s). Paul Eckman argued that the facial expression features of neutrality, happiness, sadness, and fear are primarily shown around the eyes, eyebrows, nose, and mouth.

Table 4: Features of facial expressions [25]

Facial expression	Characteristics
Anger	• Eyebrows pulled down and together • Eyes opened wide, staring hard • Lips pressed tightly together
Contempt	• Tightened and raised lip corner on one side of the face
Disgust	• Lowered eyebrows • Wrinkling on the side and bridge of the nose • Upper lip is raised in an inverted "U"
Fear	• Lower lip raised and slightly protruding • Eyebrows raised and pulled together • Raised upper eyelids • Tensed lower eyelids
Happiness	• Jaw dropped open and lips stretched horizontally backwards • Eyes are narrowed and there is some winkling around the eyes • Cheeks are raised • Lip are pulled back and teeth are exposed in a smile
Sadness	• Inner corners of the eyebrows pulled up and together • Upper eyelids drooped and eyes looking down • Lip corners pulled downward

(Continued)

Table 4 (continued)

Facial expression	Characteristics
Surprise	• Eyebrows raised, but not drawn together • Upper eyelids raised, lower eyelids neutral • Jaw dropped down

Next, we extract the landmarks [24] from the facial region as red-filled circles in Fig. 6b. Among the red-filled circles, we filter out a small number of the most relevant landmarks to facial expressions, connected with yellow lines in Fig. 6b. They are positioned around the eye, eyebrows, nose, and mouth. Finally, we can construct the feature vectors by measuring the distance between the yellow-lined circles. As in Fig. 6a, the d1-10 from the eye and eyebrows, d11-12 from the nose, and d13-18 from the mouth are the distances, each resulting in a feature vector.

Consequently, we are given feature vectors for local characteristics from the landmarks of three different regions as well as the feature vectors from the whole face area for global characteristics. Those four types of feature vectors are employed to train the respective CNN. Subsequently, the soft voting method decides on the recognized facial emotion.

As shown in Fig. 6, the landmarks expressing feature information in the proposed algorithm are related to the eyes, eyebrows, nose, and mouth among the 68 extracted landmarks as the red-filled circles in Fig. 6b. We chose them according to the facial expression-based features [25] presented in Table 4, which were suggested by Paul Eckman, who demonstrated the universality of seven primary facial expressions of emotions found in social and psychological research for several decades (since the 1970s). Paul Eckman argued that the facial expression features of neutrality, happiness, sadness, and fear are primarily shown around the eyes, eyebrows, nose, and mouth. Accordingly, as shown in Table 4, the features of a facial expression—i.e., whether the eyebrows are raised, the eyes are enlarged, or the mouth is opened—are represented as feature vectors by extracting the distance between the landmarks in Fig. 6.

After recognizing the emotions according to the feature vectors extracted from the whole facial image and the selected landmarks, as shown in Fig. 7, the ensemble learning via the soft-voting method determines the emotion with the highest probability. The CNN network consists of three convolution layers, and the size of each layer is $3 \times 3 \times 8$, $3 \times 3 \times 16$, and $3 \times 3 \times 32$, respectively, as shown in Fig. 4. The structure of the soft-voting process is shown in Fig. 7. The feature vectors from the entire face image and the distances between landmarks related to eyes and eyebrows, the distance between landmarks related to the nose, and the distance between landmarks related to the mouth consist of the training data. Each training data train the CNN networks (Classifier 1 to Classifier 4). The average probability is calculated for each class after extracting the classes of the highest and second-highest probabilities during the training process. The class with the highest average value is finally determined as the recognized emotion [26].

For example, in Fig. 7, the probability that the facial emotion recognition result will be Class 1 is 0.7 in Classifier 1, 0.2 in Classifier 2, 0.8 in Classifier 3, and 0.9 in Classifier 4. An average of 0.65 obtained by adding up all the probabilities for Class 1 indicated by these four classifiers is the final facial emotion recognition result for Class 1. The probability of becoming Class 2 is 0.3 in Classifier 1, 0.8 in Classifier 2, 0.2 in Classifier 3, and 0.1 in Classifier 4. The average of 0.35 obtained by adding up

all the probabilities for Class 2 indicated by these four classifiers is the final facial emotion recognition result for Class 2. Accordingly, Class 1, with the highest average probability value, becomes the final facial emotion recognition result. In this way, information about the distance between landmarks related to the entire face image, eyes, and eyebrows, between landmarks related to the nose, and between landmarks related to the mouth is learned using the assigned CNN network. Finally, facial expression using the soft voting method recognizes the resulting data as a specific emotion.

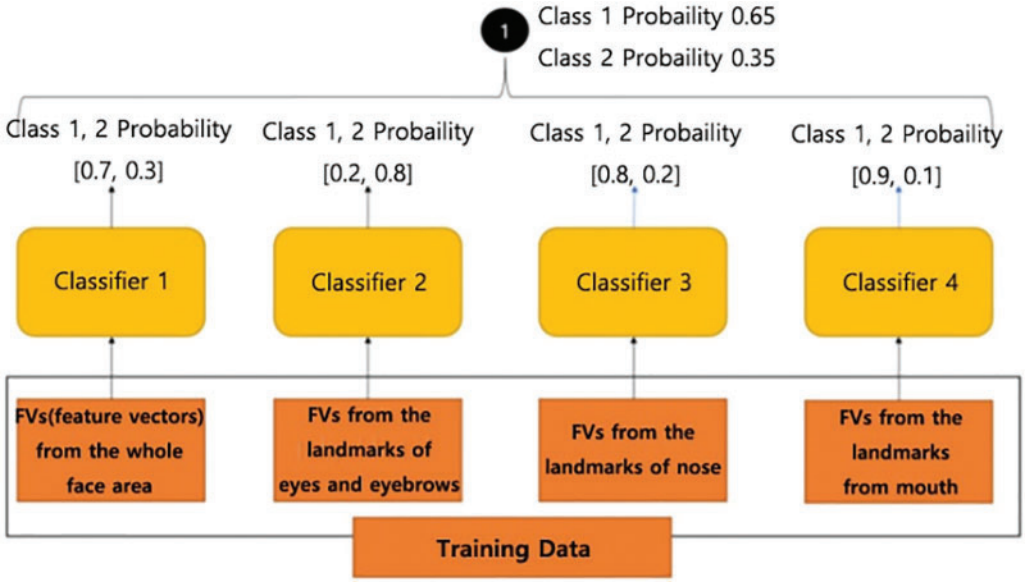


Figure 7: Soft-voting ensemble learning method

4 Experiments and Analysis

This paper proposes an emotion recognition method that extracts feature information from only specific parts of the face to overcome the weakness that facial emotion recognition methods are sensitive to illumination conditions and backgrounds. We chose three public databases to experiment with the proposed and conventional emotion recognition methods mentioned in Section 2. Finally, we comparatively analyzed the results. The experimental setup comprised a 32TFLOPS (64-bit) GPU, a 2.99-GHz 24-core CPU, 350 GB of RAM, and 200-GB SSD, and MATLAB was used. During the experiments, each database was divided into 7:3 for the training and test data ratio.

As shown in Table 5, the experiment's results for comparing the proposed algorithm with the conventional algorithms using the CK+ database are as follows: the accuracy of the proposed algorithm was 2.87% higher than that of the algorithm proposed by Mollahosseini et al. [10], which used a newly constructed deep neural network architecture, 3.6% higher than the algorithm proposed by Liu et al. [11], which detected particular facial action parts and used an adaptive 3D CNN employing discriminatory features, and 1.98% higher than that of the algorithm proposed by Happy et al. [17], which extracted landmarks from the face and used the prominent feature information among them as patch feature information, and 2.83% higher than that of the algorithm proposed by Jain et al. [18], which is a method for classifying six face graph classes based on a single deep convolutional neural network (DNN) containing convolutional layers and deep residual blocks. However, it was 0.39%

lower than the algorithm proposed by Liand et al. [12] based on DBM-CNN, in which multi-label cross-entropy losses and bi-manifold loss were added to produce deep layers.

Table 5: Comparison of the emotion recognition accuracy (%) between the proposed algorithm and the conventional algorithms for different databases

Algorithm	Database		
	CK+	JAFFE	FER2013
Proposed algorithm (Deep facial emotion recognition using global and local characteristics by exploiting facial landmarks)	96.07	96.97	95.87
Going deeper in facial-expression recognition using deep neural networks [10]	93.20	–	66.40
Deeply learning deformable facial action parts model for dynamic expression analysis [11]	92.40	–	–
Blended emotion in-the-wild: multi-label facial expression recognition using crowdsourced annotations and deep locality feature learning [12]	96.46	–	–
Deep learning using linear support vector machines [13]	–	–	69.30
Deep-Emotion: facial expression recognition using attentional convolutional network [14]	–	92.80	70.02
Facial emotion recognition: state of the art performance on FER2013 [15]	–	–	73.28
A neural network based facial expression recognition using fisherface [16]	–	89.20	–
Automatic facial-expression recognition using features of salient facial patches [17]	94.09	91.80	–
Extended deep neural network for facial emotion recognition [18]	93.24	95.23	–
Image augmentation for classifying facial expression images by using deep neural network pre-trained with object image database [19]	–	95.31	–
Hybrid deep neural networks for face emotion recognition [20]	–	94.91	–

The results of the experiment in which the proposed algorithm and the conventional algorithms were compared using the JAFFE database are as follows: the accuracy of the proposed algorithm was 4.17% higher than that of a deep-learning method proposed by Minaee et al. [14], which is based on attention convolutional networks and a technique of detecting important feature parts reflecting facial

expressions, 7.77% higher than the method proposed by Abidin et al. [16], which used a neural network-based facial expression recognition employing Fisherface features to recognize facial expressions, 5.17% higher than the method proposed by Happy et al. [17], which extracted landmarks from the face and used the prominent feature information among them as patch feature information, 1.74% higher than the algorithm proposed by Jain et al. [18], which is a method for classifying six face graph classes based on a single deep convolutional neural network (DNN) containing convolutional layers and deep residual blocks, 1.66% higher than the system proposed by Shima et al. [19], which extracted facial feature information based on deep neural networks and an SVM, and 2.06% higher than the system proposed by Jain et al. [20], which proposed a hybrid convolution-recurrent neural network.

In particular, a significant contribution of the proposed method is demonstrated in the following investigation. The results of the experiment in which the proposed algorithm and the conventional algorithms were compared using the FER2013 database were as follows: the accuracy of the proposed algorithm was 26.57% higher than that of the CNN adopted by Tang [13], which was similar to AlexNet but had top layers of linear 1-vs-all and linear SVM instead of the softmax function, 25.85% higher than that of a deep-learning method proposed by Minaee et al. [14], which is based on attention convolutional networks and a technique of detecting important feature parts showing facial expressions, and 22.59% higher than that of the method proposed by Khaireddin et al. [15], which used a VGGNet architecture with fine-tuning of the hyperparameters.

The experiments conducted using the three public databases indicated that the proposed method outperformed the conventional methods using facial expression-related feature information or reconstructed neural networks. In particular, the proposed algorithm achieved a significantly higher recognition accuracy for the FER2013 database containing various illumination conditions and backgrounds than the conventional ones.

Fig. 8 shows examples of the emotion recognition results using the FER2013 database. Figs. 8a–8d show the original images and facial expressions with illumination at different brightness levels. Figs. 8a and 8d present images captured under the basic brightness, and Figs. 8b and 8c present images with different brightness depending on the subregions in the face. We used images with varying brightness levels to evaluate facial emotion recognition accuracy. The results are as follows: facial emotion recognition accuracy was higher for the proposed algorithm than conventional algorithms, regardless of the illumination conditions. Consequently, the proposed method proved robust to illumination conditions because the feature information employed is tolerant to the change in brightness.

Fig. 8e, the background is different, and the face is tilted. In this case, the proposed method still achieved higher accuracy than the conventional methods, indicating that it is robust to the backgrounds and situations.

In addition, the emotion recognition accuracy in Figs. 8b and 8d showed that the emotion of fear was significantly higher than that of other emotions, and, in (c) and (f) of Fig. 8, the emotion of anger was significantly higher than that of other emotions. The emotion that appears before a person engages in illegal or prohibited actions is similar to fear or anger. Thus, we believe the proposed algorithm will be able to alert or prevent such criminal behaviors in advance.

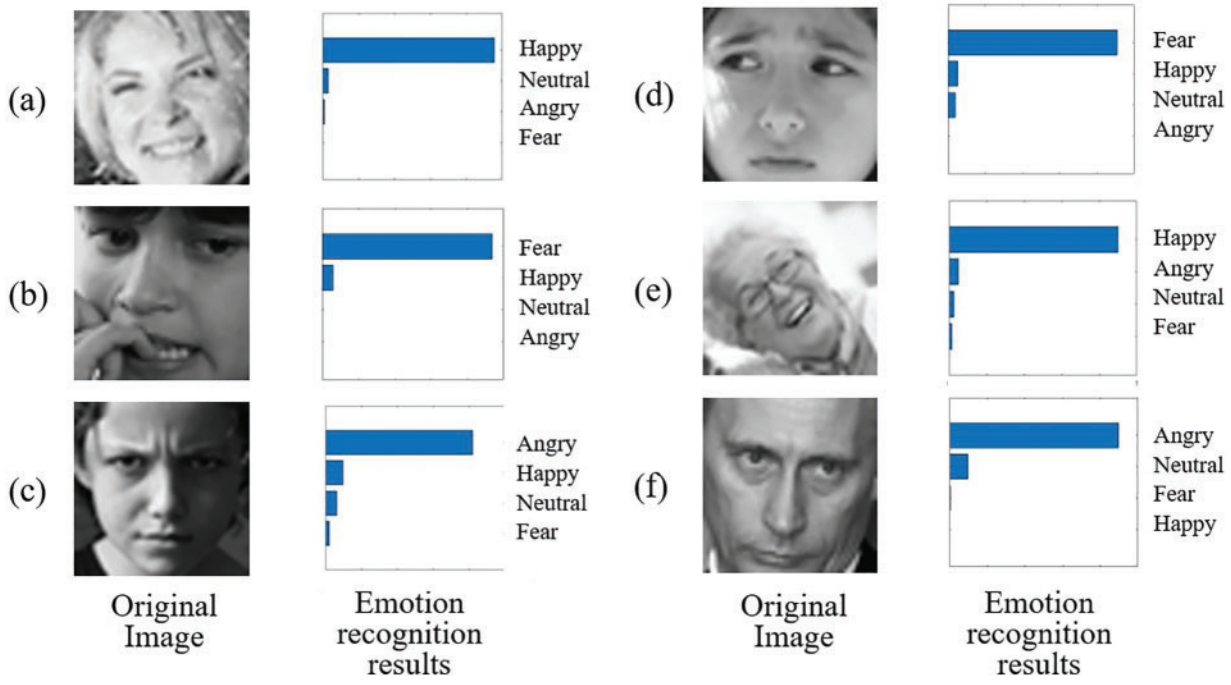


Figure 8: Facial emotion recognition results obtained using original face images with various illumination conditions and backgrounds in the FER2013 database

In order to verify that the proposed method is robust to the illumination and background conditions, we additionally executed the following experiment. We compared the performance of the proposed method with that of a method that trained only a CNN model with a whole facial image, as shown in Fig. 9. The CNN model with a whole facial image showed a performance of 72.13%, which is almost similar to the performances of the previous methods in Table 5 using the FER2013 database. However, the proposed method with feature vectors tolerant to the variation of illumination and background significantly improved up to ninety percent, which is even better than the performances of the previous methods with the CK+ and JAFFE. In other words, the proposed method using FER2013 performs as well as the previous methods using CK+ and JAFFE. It shows experimental evidence of how the feature vectors work effectively in the proposed method. In detail, as in Table 6, the accuracy of the proposed method employing the feature information from the landmarks of the face was 95.87%, i.e., 23.74% higher than that of the CNN model only using whole facial images.

According to the comparative results obtained using the CK+ and JAFFE databases and the detailed results obtained using the FER2013 database, the proposed algorithm achieved a higher facial emotion recognition accuracy than the conventional algorithms. Notably, it was confirmed that extracting features from the facial landmarks enabled the proposed method to be robust to the changes in illumination and background.

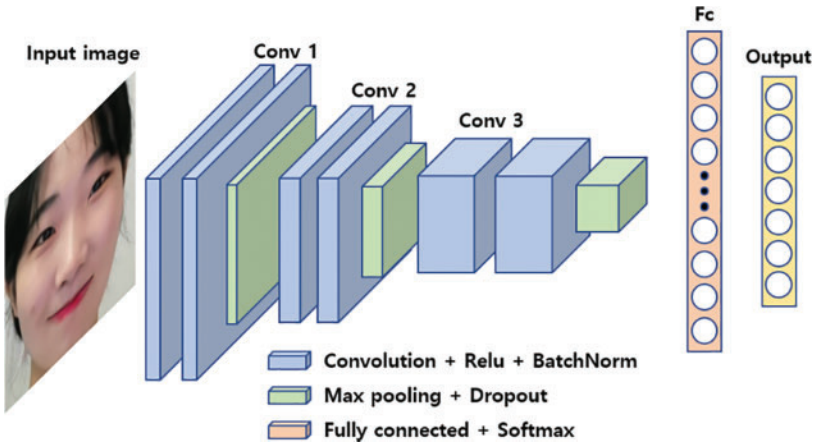


Figure 9: Sequence diagram of the training process using only a whole face image

Table 6: Comparison of facial emotion recognition accuracy between the whole facial image-based algorithm and the proposed algorithm

Algorithm	Accuracy (%)
Emotion recognition algorithm using only the whole facial image	72.13
Proposed emotion recognition algorithm	95.87

5 Conclusion

We proposed a novel emotion recognition method robust to the changes in illumination and background, which are the most challenging for recognizing emotions through facial expressions. We used the three public databases to compare the proposed and conventional facial emotion recognition methods. The proposed algorithm constructs new feature vectors by analyzing the distance between landmarks related to the eyes, eyebrows, nose, and mouth, containing numerous facial expression feature information. We used both the whole facial images and the new feature vectors to train the assigned CNNs, an ensemble network recognizing the emotion.

Public databases—CK+, JAFFE, and FER2013—were used to compare the proposed and the conventional emotion recognition methods based on facial expressions. The results indicated that the proposed method has a higher recognition accuracy than the conventional methods. Remarkably, it exhibited robustness against the changes in illumination and background, which are severe constraints when extracting facial expression-related features. The experimental results using the FER2013 database, which contains facial expression data with various illumination conditions and backgrounds, demonstrated such robustness. Therefore, we believe the proposed method can be used in various fields, including psychology, neurology, behavioral science, and computer science. In particular, concerning security, facial expression analysis can detect dangerous situations such as terrorism, robbery, and fighting in advance to prevent major accidents. Such an application could extend the boundaries of human-computer interaction. To this end, future research should improve emotion recognition performance by studying multi-information-based emotional state classification using detailed feature

pattern information, such as facial movement feature extraction, and catching facial expression changes in continuous or dynamic images instead of a single image.

Acknowledgement: This work was supported by Artificial intelligence industrial convergence cluster development project funded by the Ministry of Science and ICT (MSIT, Korea) & Gwangju Metropolitan City.

Funding Statement: This research was supported by the Healthcare AI Convergence R&D Program through the National IT Industry Promotion Agency of Korea (NIPA) funded by the Ministry of Science and ICT (No. S0102-23-1007) and the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2017R1A6A1A03015496).

Author Contributions: Conceptualization, Y.-E.A., J.-M.L. and E.B.; Methodology, Y.-E.A. and E.B.; Software, Y.-E.A. and J.-M.L.; Validation, Y.-E.A., E.B. and S.P.; Formal analysis, Y.-E.A.; Investigation, Y.-E.A.; Writing—Original Draft Preparation, Y.-E.A.; Writing—Review and Editing Y.-E.A., J.-M.L., E.B. and S.P.; Supervision, S.P.; Project Administration, S.P.; Funding Acquisition, S.P. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] K. M. Huh and S. M. Kang, "Facial expression recognition technology," *Journal of Institute of Control, Robotics and Systems, Institute of Control Robotics and Systems*, vol. 20, no. 2, pp. 39–45, 2014.

[2] A. A. M. Al-Modwahi, O. Sebetel, L. N. Batleng, B. Parhizkar and A. H. Lashkari, "Facial expression recognition intelligent security system for real time surveillance," in *World Congress in Computer Science, Computer Engineering, and Applied Computing (WORLDCOMP'12)*. Las Vegas, Nevada, United States of America, pp. 1–8, 2012.

[3] M. Sajjad, M. Nasir, F. U. M. Ullah, K. Muhammad, A. K. Sangaiah *et al.,* "Raspberry Pi assisted facial expression recognition framework for smart security in law-enforcement services," *Journal of the Information Sciences*, vol. 479, pp. 416–431, 2019.

[4] K. H. Kim, K. Park, H. Kim, B. Jo, S. H. Ahn *et al.,* "Facial expression monitoring system for predicting patient's sudden movement during radiotherapy using deep learning," *Journal of Aapplied Clinical Medical Physics*, vol. 21, no. 8, pp. 191–199, 2020.

[5] Y. H. Kim, J. Y. Kim and K. T. Bae, "A study on artificial intelligence-based intensive medical patient's shock pre-detection system," *Journal of Information Technology and Applied Engineering*, vol. 9, no. 2, pp. 49–56, 2019.

[6] J. H. Han and C. S. Chung, "Mapping facial expressions onto internal states," *Korean Journal of the Science of Emotion & Sensibility, Korean Society for Emotion and Sensibility*, vol. 1, no. 1, pp. 41–58, 1998.

[7] J. A. Russel and B. Fehr, "Relativity in the perception of emotion in facial expressions," *Journal of Personality and Social Psychology*, vol. 116, no. 3, pp. 223–237, 1987.

[8] G. Rhodes, "Looking at faces: First-order and second-order features as determinants of facial appearance," *Perception*, vol. 17, no. 1, pp. 43–63, 1988.

[9] I. Pilowsky, M. Thornton and B. Stokes, "Towards the quantification of facial expressions with the use of a mathematic model of the face," in *Aspects of Face Processing, NATO ASI Series,* vol. 28. Berlin: Springer, pp. 340–348, 1986. [Online]. Available: https://doi.org/10.1007/978-94-009-4420-6_36

[10] A. Mollahosseini, D. Chan and M. H. Mahoor, "Going deeper in facial expression recognition using deep neural networks," in *2016 IEEE Winter Conf. on Applications of Computer Vision (WACV)*, Lake Placid, NY, USA, IEEE, pp. 1–10, 2016.
[11] M. Liu, S. Li, S. Shan, R. Wang and X. Chen, "Deeply learning deformable facial action parts model for dynamic expression analysis," in *Computer Vision-ACCV 2014*. Singapore: Springer, pp. 143–157, 2014.
[12] S. Liand and W. Deng, "Blended emotion in-the-wild: Multi-label facial expression recognition using crowdsourced annotations and deep locality feature learning," *International Journal of Computer Vision*, vol. 127, no. 6–7, pp. 884–906, 2019.
[13] Y. Tang, "Deep learning using linear support vector machines," *Contribution to the ICML 2013 Challenges in Representation Learning Workshop*, 2015. https://doi.org/10.48550/arXiv.1306.0239
[14] S. Minaee and A. Abdolrashidi, "Deep-Emotion: Facial expression recognition using attentional convolutional network," *Sensors*, vol. 21, no. 9, pp. 1–16, 2021.
[15] Y. Khaireddin and Z. Chen, "Facial emotion recognition state of the art performance on FER2013," *Computer Vision and Pattern Recognition*, 2021. https://doi.org/10.48550/arXiv.2105.03588
[16] Z. Abidin and A. Harjoko, "A neural network based facial expression recognition using fisherface," *International Journal of Computer Applications*, vol. 59, no. 3, pp. 30–34, 2012.
[17] S. L. Happy and A. Routray, "Automatic facial expression recognition using features of salient facial patches," *IEEE Transactions on Affective Computing*, vol. 6, no. 1, pp. 1–12, 2015.
[18] D. K. Jain, P. Shamsolmoali and P. Sehdev, "Extended deep neural network for facial emotion recognition," *Pattern Recognition Letters*, vol. 120, pp. 69–74, 2019.
[19] Y. Shima and Y. Omori, "Image augmentation for classifying facial expression images by using deep neural network pre-trained with object image database," in *Proc. of the 3rd Int. Conf. on Robotics, Control and Automation*, Chengdu, China, ACM, pp. 140–146, 2018.
[20] N. Jain, S. Kumar, A. Kumar, P. Shamsolmoali and M. Zareapoor, "Hybrid deep neural networks for face emotion recognition," *Pattern Recognition Letters*, vol. 115, no. 2, pp. 101–106, 2018.
[21] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. of the 2001 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR)*, Kauai, HI, USA, pp. 511–518, 2001.
[22] V. Kazemi and J. Sullivan, "One millisecond face alignment with an ensemble of regression trees," in *2014 IEEE Conf. on Computer Vision and Pattern Recognition*, Columbus, OH, USA, pp. 1867–1874, 2014.
[23] J. H. Kim and C. S. Won, "Emotion enhancement for facial images using GAN," in *2020 IEEE Int. Conf. on Consumer Electronics—Asia (ICCE-Asia)*, Seoul, Korea (South), pp. 1–4, 2020.
[24] B. T. Nguyen, M. H. Trinh, T. V. Phan and H. D. Nguyen, "An efficient real-time emotion detection using camera and facial landmarks," in *Seventh Int. Conf. on Information Science and Technology*, Da Nang, Vietnam, pp. 251–255, 2017.
[25] P. Ekman and W. Friesen, "Facial action coding system: A technique for the measurement of facial movement," Palo Alto: Consulting Psychologists Press, 1978. [Online]. Available: https://www.paulekman.com/facial-action-coding-system/
[26] S. W. A. Sherazi, J. W. Bae and J. Y. Lee, "A soft voting ensemble classifier for early prediction and diagnosis of occurrences of major adverse cardiovascular events for STEMI and NSTEMI during 2-year follow-up in patients with acute coronary syndrome," *PLoS One*, vol. 16, no. 6, pp. 1–20, 2021.