Feature Extraction Techniques for Human Emotion Identification from Face Images

M. Kalpana Chowdary, D.Jude Hemanth, Angelopoulou, A. and Kapetanios, E.

This is a copy of the author's accepted version of a paper subsequently published in the proceedings of the 9th International Conference on Imaging for Crime Detection and Prevention, 16 - 18 Dec 2019. doi:10.1049/cp.2019.1173 .

The final published version will be available online at:

https://dx.doi.org/10.1049/cp.2019.1173

Feature Extraction Techniques for Human Emotion Identification from Face Images

M. Kalpana Chowdary[1], D.Jude Hemanth[1,*], A. Angelopoulou[2], E. Kapetanios[2]

[1]Department of ECE, Karunya Institute of Technology and Sciences, Coimbatore, India
[2]Cognitive Computing Research Lab, University of Westminster, UK
*Corresponding Author: judehemanth@karunya.edu

Keywords: Emotion Recognition, Feature extraction, MSER, Connected Components, ANN.

Abstract

Emotion recognition has been one of the stimulating issues over the years due to the irregularities in the complexity of models and unpredictability between expression categories. So many Emotion detection algorithms have developed in the last two decades and still facing problems in accuracy, complexity and real-world implementation. In this paper, we propose two feature extraction techniques: Mouth region-based feature extraction and Maximally Stable Extremal Regions (MSER) method. In Mouth based feature extraction method mouth area is calculated and based on that value the emotions are classified. In the MSER method, the features are extracted by using connecting components and then the extracted features are given to a simple ANN for classification. Experimental results shows that the Mouth area based feature extraction method gives 86% accuracy and MSER based feature extraction method outperforms it by achieving 89% accuracy on DEAP. Thus, it can be concluded that the proposed methods can be effectively used for emotion detection.

1. Introduction

Emotions are synonymous with human behaviour and play an important role in influencing the way we think and behave. Emotions enable humans to perceive how they feel about something [1]. Emotion recognition can be determined by using facial expressions, body gestures, speech, and Physiological signals. Emotions and gestures are forms of non-verbal communication and voice is a form of verbal communication. Face Emotion recognition is one of the most significant issues used in the applications of Human-machine interaction, Drowsiness detection in driving, Counselling systems, entertainment, robotics, surveillance, health care, psychology study and many more [2] [3] . In [4], the author analyses the applications which could be easy to use and are user-friendly, that even people with specific disabilities can use easily. In marketing customers, satisfaction/dissatisfaction with the product can be estimated by observing facial expressions [5]. One of another advantage of automatic facial expression recognition in the medical field is that it helps the doctors to identify the patient's mental health or medicine reaction in no or less time. Another advantage in the educational field is obtaining teaching feedback based on the facial expression identification of the students attending the particular class. Facial emotion identification has major roles and challenges in the areas of computer vision, pattern recognition which has gained greater attention due to potential applications in many fields. Now a day's emotion recognition is one of the most demanding technology. Facial expressions based on human- machine analysis have become one of the popular research areas. To effectively and automatically classify human emotions based on facial expressions, it is necessary to initially detect the faces followed by feature extraction and finally, classification based on the extracted features.

According to Ekman[6], the emotions are classified into six categories namely happy, sad, surprise , angry, disgust and fear based on the facial action coding system (FACS) that consists of action units (AUs) using facial movement. Many researchers have used AUs in their recognition systems to classify the expressions and in some kinds of literature facial animation parameters are also considered. The author of [7] gives a detailed review of conventional and deep learning-based emotion recognition approaches and also has provided a list of facial emotion databases with web links. The research paper [8] gives the detailed survey on all existing pre-processing, feature extraction and classification techniques and listed the performance analysis of various facial emotion recognition approaches. The author stated that ROI segmentation in pre-processing gives a high accuracy of 99% and in feature extraction, Gabor filter gives a high accuracy of 85-99% with less computational time and finally, proved that among all classifiers Support Vector Machines (SVM) gave highest recognition rate.

The paper mainly concentrates on classifying the emotion with respectively pre-processing, feature extraction, Feature selection, and classification techniques. This paper is further subdivided into the following mentioned sections. Section 2 discusses the problem statement, section 3 describes the flow of Emotion recognition system and information about the database, section 4 explains feature extraction techniques, section 5 discusses artificial neural networks, section 6 shows results of our system, section 7 is comparisons and final section is the conclusion.

2. Problem Statement

In the existing literature, emotion recognition based on facial expressions does not take into account the user personality and current situation of the person [9]. Another drawback faced in facial recognition techniques is the absence of temporal analysis for facial expressions to distinguish between emotions. Some of the other problems identified in the literature are different illuminations, occlusions, and pose variations in images[10]. Finally, from the analysed existing systems [11], [12] classification of emotion states also provided a low accuracy. By considering all the above limitations our main intention is to develop an emotion recognition with better accuracy.

3. Facial Emotion Recognition System

The formation of any emotion recognition systems follows mainly four steps: pre-processing, feature extraction, feature selection and classification [13] as shown in below figure 1.

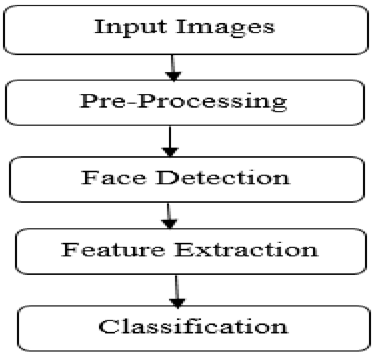

Figure 1: Emotion Recognition System

The input images are taken from the databases/real environments, after that the images are pre-processed to remove the noises and redundant information. And some of the pre-processing technique involves cropping, pixel brightness transformation, and segmentation, etc. The features are extracted from the interest areas such as eyes, eyebrows nose and mouth. Feature selection is used to reduce the complexity by eliminating unwanted features. These features are then given to classifier for particular emotion classification.

3.1 Facial Expression Database

JAFFE (Japanese Female Facial Expressions) database was developed by Michael Lyons, Miyuki Kamachi, and Jiro Gyoba [14]. The details of the database are listed in table 1, and sample images from the database are shown in figure 2.

Total number of images	213
Number of subjects	10
Number of Emotions	7 (Happy, Fear, Sad, Surprise, Disgust, Neutral and Angry)
Color information	grey
size	256*256

Table 1: Information of JAFFE Database

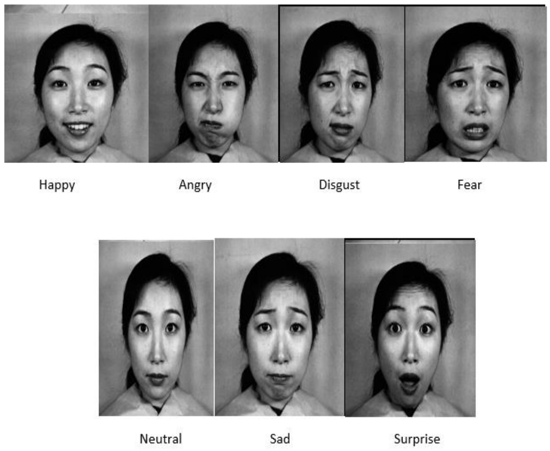

Figure 2: Sample Facial expression from JAFEE Database

3.2 Image pre-processing

Image databases contain different factors such as noise, blurring, etc. which alters the results of the analysis. To obtain results with maximum accuracy pre-processing is initially done on the images. The image pre-processing procedure is performed by following certain steps which are image enhancement, image resizing and noise reduction. Various pre-processing techniques are designed for noise removal in images like flipping, cropping [15], sharpening etc. Smoothing is also one of the effective methods to remove impulse noise in an image. It reduces the high-frequency components in an image and retains low-frequency components to smooth the image. In this work, cropping is used as a pre-processing step. The advantages of image cropping are reduced computation time, less memory for storage and recognition speed also increased. The original images of size 256 × 256 are cropped into the size of 164×164 as shown in the below figures.

Figure 3: a) Input Images b) Pre Processed Images

4. Feature Extraction Techniques

The fundamental step in the emotion recognition process is feature extraction. A set of features that are extracted from the facial images are given as input to the classifier and the accuracy of the classification depends on these features.

4.1 Mouth Feature Extraction Method

Emotions can be detected by extracting the area of the mouth regions. The main steps involved in Mouth area calculation are mouth region segmentation, conversion of the grey/color segmented mouth region into binary and finally, morphological operations are performed to obtain the mouth region then area is calculated [16], [17]. The schematic diagram for extracting the mouth region is shown below.

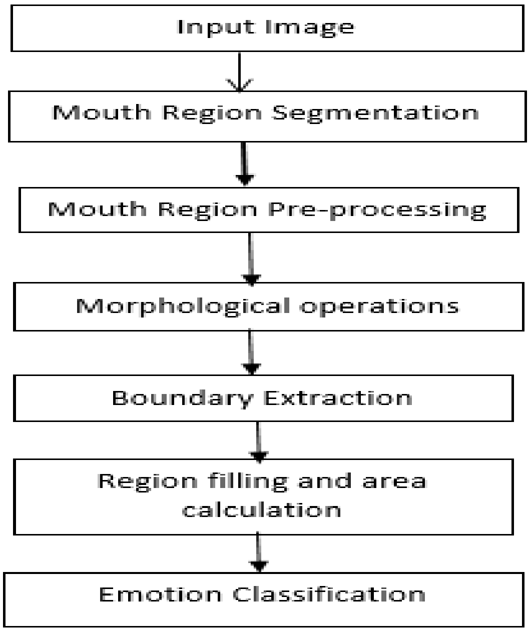

Figure 4: Schematic diagram for Mouth area calculation

In the first step, input images are selected from the Database, and mouth region segmentation is performed by using the Viola-Jones Mouth detection algorithm [18]. After separating the mouth region pre-processing is performed to convert the grey color mouth regions into binary. In the next step the binary image is complemented[19], [20].

The morphological opening function is used to preserve the shape of the large objects and also to remove the small objects from an image. The function imopen ( ) with a valued structuring element is used in the morphological opening process.

SE = strel(‘disk’,12); (1)

opening = imopen(image, SE); (2)

And next imerode( ) function is applied to the opened image to shrink the foreground in the object.

SE1= strel(‘disk’, 2,0); (3)

Erosion = imerode(opening, SE1); (4)

Boundary extraction is obtained by the subtraction of opening image from Erosion image.

BE= Erosion- opening (5)

In the next step bwareaopen( ) function is applied to extract only the mouth region.

a= bwareaopen(BE, 80); (6)

Finally, imfill( ) function is applied to fill any holes in the mouth region.

b= imfill(a, holes); (7)

The area of the mouth region is calculated by using the following formula.

Aarea = sum(b(:)); (8)

Area = Aarea * 0.2645; (9)

The below figure 5 shows the results of the intermediate stages.

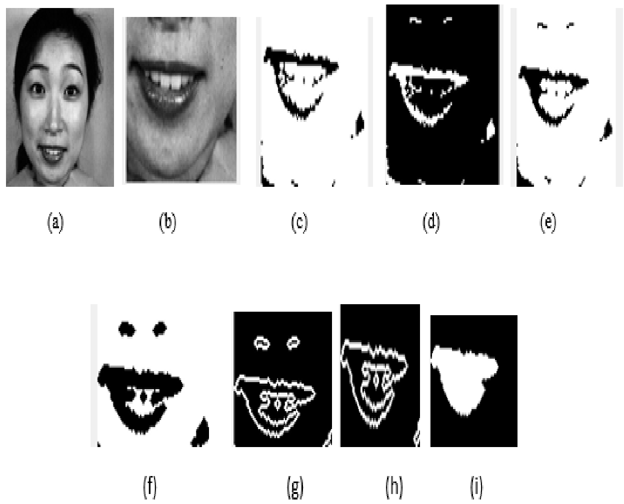

Figure 5: (a) Input image (b) Mouth region segmentation (c) Binary (d) Complemented (e) Opening (f) Erosion (g) Boundary extraction (h) Area open (i) Holes filled

Based on the area values the emotions are classified into happy, fear, sad, angry and surprise.

4.2 Feature Extraction using MSER

The theory of Maximally Stable Extremal Regions was introduced by Matas et al [21]. The initial step in MSER is to take threshold values from 0 to 255 and get the corresponding binary images. After obtaining all the binary images, some connected regions are unchanged with varying intensity values which are called the maximally stable extremal regions. The position of the stable regions and the respective threshold values are taken as MSER key points. The mathematical equation is shown below

$$q(i) = |Q_{i+\Delta} \setminus Q_{i-\Delta}| / |Q_i| \quad (10)$$

Where Q_i : connected region of threshold i

Δ : Change of grey value

$q(i)$: Q_i region changing rate with threshold i

When q(i) is the local minimal , then the Q_i is considered as the maximally stable extremal regions.

The implementation of MSER algorithm [22] follows the below steps.

Step 1: Input images from the database

Step 2: Intensity-based segmentation

Step 3: Finding the connected components and labelling the connected components

Step 4: filter the image to suppress gradients near edges

Step 5: set new threshold value to supress unwanted corners

Step 6: Take the maximum value in each component and also consider X &Y coordinates of interest points

Step 7: Finally classification is based on extracted features

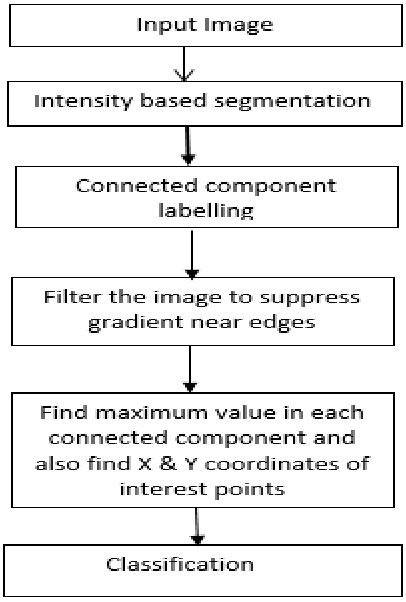

Figure 5: MSER algorithm

5. Classification Using ANN

The usage of ANN is increasing day by day in the fields of pattern matching, classification, data mining, gaming, bioinformatics, and robotics. The main layers involved in the design of neural networks are the input layer, hidden layer, and output layer [23],[24]. The processing of data is done in hidden layers depending upon the weights attached to the nodes. Weights are adjusted to minimize the error between inputs and targets. Many learning rules have been implemented to modify the weights according to the inputs. The learning rules are categorized into three types: supervised learning, unsupervised learning and reinforcement learning. Among all the learning rules, Backpropagation under supervised learning is most preferable and widely used. The architecture of simple ANN is shown in the figure below.

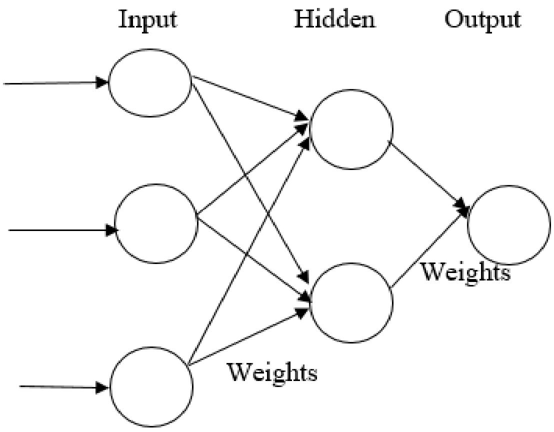

Figure 6**:** Simple ANN architecture

The training process of the above mentioned neural network is given in [25].

6. Experimental Results

In this work, a total of fifty images from the JAFFE database were considered to find emotions such as neutral, happy, sad and surprise. The performance metrics used to evaluate the proposed work are accuracy, sensitivity, and specificity.

Accuracy – It is defined as the ratio of correct evaluations to the total number of evaluations, it is the total number of correct retrieved output images from the image dataset. It shows the quality of the algorithm in retrieving images from a dataset.

$$\text{Accuracy} = \frac{TN+TP}{TN+TP+FN+FP} \tag{11}$$

Where TN represents True Negative, TP (True Positive), FN (False Negative), and FP (False Positive)

Sensitivity – It is the value obtained as the ratio of true positive evaluation to the summation of true positive value and false negative values. It is always measured in percentage.

$$\text{Sensitivity} = \frac{TP}{TP+FN} \tag{12}$$

Specificity – The number of true negative evaluation to the summation number of false-positive and true negative evaluation, it is measured in percentage.

$$\text{Specificity} = \frac{TN}{TN+FP} \tag{13}$$

The experimental results obtained based on mouth region feature extraction and MSER methods are given below.

6.1 Results of Mouth Region-based Feature Extraction

The detected mouth regions are shown in the below figures. By considering the white pixels count, area value is calculated.

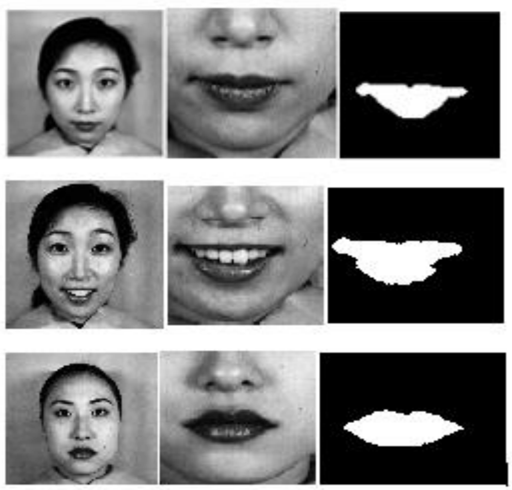

Figure 7: a) Input images b) Mouth Segmentation c) Mouth area

Table 2 shows a different range of area values for different emotions of Neutral, Sad, Surprise and Happy.

Facial Emotions	Mouth area range	
	Min value	Max value
Neutral	130	190
Sad	190	230
Surprise	230	280
Happy	260	360 (some cases overlapping with surprise)

Table 2: Mouth area values for different Emotions

Table 2 shows different mouth area ranges in that the range of 130 to 190 for neutral, 190 to 230 for sad, 230 to 280 surprise and 260 to 360 for happy.

		Predicted Emotions			
		Happy	Neutral	Sad	Surprise
Actual Emotions	Happy	7	0	0	1
	Neutral	0	12	4	1
	Sad	4	1	7	1
	Surprise	1	0	1	10

Table 3: Confusion matrix of the classifier

SE= Sensitivity, SP = Specificity, A= Accuracy

	TP	TN	FP	FN	SE	SP	A
Happy	7	37	5	1	0.87	0.88	88
Neutral	12	32	1	5	0.70	0.96	90
Sad	7	32	5	6	0.53	0.86	78
Surprise	10	35	2	3	0.76	0.92	90
Average Results					0.71	0.90	86.5%

Table 4: Performance measures of mouth area based emotion recognition

The experimental results show that the overall recognition rate of our proposed method is 86%. Among all four emotions, neutral and surprise achieve the highest recognition rate of 90%. For happy and sad the recognition rates are 88% and 78% respectively.

6.2 Results of MSER descriptor based Feature Extraction

From the given input image, the face is detected and then ROI is identified to extract features using MSER. The input image is given in figure 8.

Figure 8: Input face image

The face detection is done by using a viola-jones detection algorithm and the detected face is shown in figure 9.

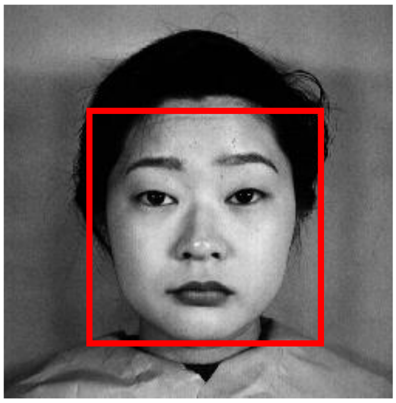

Figure 9: Face detected image using viola jones algorithm

After detecting the face, the ROI of the image is obtained by cropping the image with a fixed size.

Figure 10: Region of interest after detecting face

Finally, Feature extracted image using MSER as shown in figure11. Below figure shows the connected components in the image.

Figure 11: Connected components of the image

Key points are extracted using the MSER method. The first feature is the maximum value of the connected component and the remaining two features are the coordinates of interest points. Finally, extracted MSER features are trained by using ANN and obtained an accuracy of 89 %. It is calculated by using the metrics mentioned in section 6.

7. Comparisons

Below table 5 shows comparison with existing facial emotion techniques

Work	Features	Accuracy
Our approach 2	MSER features	89%
Our approach 1	Mouth area	86%
Bharati A. Dixit et al.,[26]	Zernike moments	81.66%
Zhan Zhang et al.,[27]	Temporal and frequency features	80%
Singh et al.,[28]	Local Binary patterns	81%
Austin Nicolai et al.,[29]	Eyes , mouth & eyebrows (centre , corner points)	79%

Table 5: Comparison to state-of-the-art approaches

Four Existing techniques are compared with proposed approaches, the recognition rate is increased.

8. Conclusion

In this paper two facial feature extraction techniques are proposed Mouth area-based feature extraction method and the MSER method. The classification was performed using ANN. The performance analysis is done based on the accuracy of the system. Based on feature extraction techniques MSER provides better compared to the Mouth region-based feature extraction method. Classification of facial emotion is one of the key technologies in many applications. In the future, it is possible to implement this feature extraction techniques with other classifiers for better accuracy.

These approaches are normally used in crime detection applications. Specifically, these techniques are used to extract the truth from the culprits. The face expressions are used as an indicator of the emotional status of the culprits. These emotions combined with EEG signals are used to analyse the nature of information (true/false) revealed by the criminals.

References

[1] M. K. Mandal and A. Awasthi, Eds., *Understanding Facial Expressions in Communication*. New Delhi: Springer India, 2015.

[2] C. Gagliardi, E. Frigerio, D. M. Burt, I. Cazzaniga, D. I. Perrett, and R. Borgatti, “Facial expression recognition in Williams syndrome,” *Neuropsychologia*, vol. 41, no. 6, pp. 733–738, Jan. 2003.

[3] Zhihong Zeng, M. Pantic, G. I. Roisman, and T. S. Huang, “A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 1, pp. 39–58, Jan. 2009.

[4] N. K. Gondhi and Er. N. Kour, “A comparative analysis on various face recognition techniques,” *International Conference on Intelligent Computing and Control Systems (ICICCS)*, Madurai, 2017, pp. 8–13.

[5] L. Z. Ruiz, R. P. V. Alomia, A. D. Q. Dantis, M. J. S. S. Diego, C. F. Tindugan, and K. K. D. Serrano, “Human emotion detection through facial expressions for commercial analysis,” *IEEE International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM)*, Manila, Philippines, 2017, pp. 1–6.

[6] P. Ekman, E. R. Sorenson, and W. V. Friesen, “Pan-Cultural Elements in Facial Displays of Emotion,” *Science*, vol. 164, no. 3875, pp. 86–88, Apr. 1969.

[7] B. Ko, “A Brief Review of Facial Emotion Recognition Based on Visual Information,” *Sensors*, vol. 18, no. 2, p. 401, Jan. 2018.

[8] I. M. Revina and W. R. S. Emmanuel, “A Survey on Human Face Expression Recognition Techniques,” *J. King Saud Univ. - Comput. Inf. Sci.*, p. S1319157818303379, Sep. 2018.

[9] L. Zhang, D. Tjondronegoro, and V. Chandran, “Random Gabor based templates for facial expression recognition in images with facial occlusion,” *Neurocomputing*, vol. 145, pp. 451–464, Dec. 2014.

[10] B. Martinez and M. F. Valstar, “Advances, Challenges, and Opportunities in Automatic Facial Expression Recognition,” in *Advances in Face Detection and Facial Image Analysis*, M. Kawulok, M. E. Celebi, and B. Smolka, Eds. Cham: Springer International Publishing, 2016, pp. 63–100.

[11] M. Dahmane and J. Meunier, “Prototype-Based Modeling for Facial Expression Analysis,” *IEEE Trans. Multimed.*, vol. 16, no. 6, pp. 1574–1584, Oct. 2014.

[12] L. Zhang and D. Tjondronegoro, “Facial Expression Recognition Using Facial Movement Features,” *IEEE Trans. Affect. Comput.*, vol. 2, no. 4, pp. 219–229, Oct. 2011.

[13] M. Dubey and L. Singh, “Automatic Emotion Recognition Using Facial Expression: A Review,” vol. 03, no. 02, p. 5.

[14] "Facial Expression Database: Japanese Female Facial Expression (JAFFE) Database." [Online]. Available: http://www.kasrl.org/jaffe.html. [Accessed: 20-Aug-2019].

[15] H. Ali, M. Hariharan, S. Yaacob, and A. H. Adom, "Facial emotion recognition using empirical mode decomposition," *Expert Syst. Appl.*, vol. 42, no. 3, pp. 1261–1277, Feb. 2015.

[16] A. M. Rani and R. Durgadevi, "Image Processing Techniques To Recognize Facial Emotions," vol. 6, no. 6, p. 6.

[17] G. Kalaivani and D. K. Krishnaveni, "Facial Emotion Detection Based on the Features of Mouth Regions," 2018.

[18] Y.-Q. Wang, "An Analysis of the Viola-Jones Face Detection Algorithm," *Image Process. Line*, vol. 4, pp. 128–148, Jun. 2014.

[19] M. Prasad and A. Danti, "Classification of Human Facial Expression based on Mouth Feature using SUSAN Edge Operator," 2014.

[20] M. Prasad, M. Dileep, and A. Danti, "Human Mood Classification Based on Eyes using Susan Edges," *Int. J. Data Min. Tech. Appl.*, vol. 4, pp. 23–26, Jun. 2015.

[21] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust Wide Baseline Stereo from Maximally Stable Extremal Regions," p. 10.

[22] R. Rajeswari and P. Ramya, "Performance Evaluation of Descriptors Extracted by MSER Detector for Human Action Recognition," vol. 8, no. 2, p. 9, 2018.

[23] S. Lek and J. F. Guégan, "Artificial neural networks as a tool in ecological modelling, an introduction," *Ecol. Model.*, vol. 120, no. 2–3, pp. 65–73, Aug. 1999.

[24] A. K. Jain, Jianchang Mao, and K. M. Mohiuddin, "Artificial neural networks: a tutorial," *Computer*, vol. 29, no. 3, pp. 31–44, Mar. 1996.

[25] M. Otair and A. S. Walid, "Efficient Training of Backpropagation Neural Networks," *Neural Netw. World*, vol. 16, Jan. 2006.

[26] B. Dixit and A. Gaikwad, "Statistical moments based facial expression analysis," *IEEE Int. Adv. Comput. Conf. IACC 2015*, pp. 552–557, Jul. 2015.

[27] Z. Zhang, L. Cui, X. Liu, and T. Zhu, "Emotion Detection Using Kinect 3D Facial Points," *IEEE/WIC/ACM International Conference on Web Intelligence (WI)*, Omaha, NE, USA, 2016, pp. 407–410.

[28] S. Singh, R. Maurya, and A. Mittal, "Application of Complete Local Binary Pattern Method for facial expression recognition," 2012, pp. 1–4.

[29] A. Nicolai and A. Choi, "Facial Emotion Recognition Using Fuzzy Systems," *IEEE International Conference on Systems, Man, and Cybernetics*, Kowloon Tong, Hong Kong, 2015, pp. 2216–2221.