ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 5, No. 5, 24-28 (2020)
www.astesj.com

ASTES Journal
ISSN: 2415-6698

Facial Expression Recognition using Facial Landmarks: A novel approach

Rohith Raj S*, Pratiba D, Ramakanth Kumar P

Department of Computer Science and Engineering, RV College of Engineering, Bengaluru, 560059, India

ARTICLE INFO

Article history:
Received: 18 July, 2020
Accepted: 11 August, 2020
Online: 08 September, 2020

Keywords:
CLAHE
Facial expression recognition
Facial landmarks
SVM

ABSTRACT

The universally common mode of interaction is the human emotions. Thus, there are several advantages of automated recognition of human facial expressions. The primary objective of the proposed framework in this paper, is to classify a person's facial expression into anger, contempt, disgust, fear, happiness, sadness and surprise. Firstly, CLAHE is performed on the image and the faces are identified using a histogram of oriented gradients. Then, using a model trained with the iBUG 300-W dataset the facial landmarks are predicted. Using the proposed method with the normalized landmarks, a feature vector is calculated. With this calculated feature vector, the emotions can be recognized using a Support Vector Classifier. The Support Vector Classifier was trained and tested for system accuracy using the famous Extended Cohn-Kanade database.

1 Introduction

The primary task of Facial Expression Recognition(FER) is to classify the expressions on human face images into various categories viz., happiness, fear, anger, sadness, surprise and so on. It's a crucial component of psychology as the facial expression of a individual accounts for 55% of a spoken message's impact [1].

Psychologists have historically trained human observers to recognise changes in facial muscles and use a Facial Action Coding System (FACS) to map muscle movements to an emotion [2]. While this method helped to maintain objectivity and had the power of definition, it had many disadvantages such as: ineffective preparation of human observers, mandatory for observers with a strong background in psychology and an average person spent more than 100 hours reading the FACS manual to prepare for certification, the certification itself took about twelve hours, and once certified, FACS professionals could detect emotion with an accuracy of approximately 80%.

Automatic facial expression recognition has become an interesting and challenging area for the fields of computer vision and artificial intelligence with the advent of faster computers and the use of pixels/megapixels for image elements. As we know, the traditional application of FER is in the field of psychology, where facial expressions are used to understand behaviour, detect mental disorders, and detect lies. However, applications of computer-based FER can be broadly classified into three groups, and these are, human computer interaction (HCI), consumer products, and medical research. Human-centered systems need to be built in the field of HCI in such a way that system responds not only according to user input but also according to user behaviour, particularly in ubiquitous computing environments [3]. Automotive fatigue detection systems, entertainment systems such as gaming, human-robot interaction, and protection are a few examples of commercial applications [4]. Identification of clinical depression, pain assessment, and conduct of psychology studies, medicine, are several applications related to medical research [5].

The remainder of the paper is structured as follows. Section II focuses on related research carried out on this area. Section III deals with the proposed methodology. Section IV throws light on the experimental results and analysis. And finally, Section V concludes the paper.

2 Related Work

A lot of research work in the field of human facial expression recognition has been carried out in the recent past, mainly due to the numerous use cases present in this area.

One of the earliest attempt was made by [2] in the year 1978. They had proposed FACS for FER. It was shown that each emotion is a combination of several Action Units (AU) present in FACS. And thus, concluded that facial muscular movement maps to an emotion.

In [6], they trained a Conventional Neural Network (CNN) to perform FER. A 64*64 image is supplied as input to the CNN. The network includes one input layer, five convolution layers, three pooled layers, one fully connected layer, and one output layer. The fully connected layer is combined as the input of the softmax layer to obtain the output class after the convolutional pooling operation. They have used JAFFE and CK+ database for validation.

*Corresponding Author: Rohith Raj S, RV College of Engineering, +91-8095969630 & rohith.june6@gmail.com

In [7], they trained a CNN to perform FER by using CMU MultiPie database for validation. Their proposed method also took advantage of GPU-based parallelism to boost performance.

In [8], they put forward the fact of any FER technique contains 3 steps: feature extraction, dimensionality reduction, and classification. According to them dimensionality and feature selection are major issues concerning FER. They also said that huge amount of memory and processing is required to process the image as a whole. So, they proposed an alternative which is geometric features and have used facial landmark detection for feature extraction and CNN Classifier. They used JAFEE, MUG, CK and MMI database to test effectiveness of their system.

The maximum peak frame chosen is used for the recognition of facial expression in the system designed by [9]. Their approach was relying on the calculation of the distance between the neutral and expressive face. They used eNTERFACE database to test the effectiveness of their system and attained a prediction accuracy of 78.26%.

Using 63 facial landmark points of the active appearance model, a FER system was proposed by [10]. They used those 63 points to calculate the remaining 4 landmark points. Ratio of height to width was used to measure the degree of openness. Facial expression was obtained by multiplying the respective weights with their sum of ratios. Thus, they attained a prediction accuracy of 88%.

3 Proposed Methodology

This section describes the complete proposed methodology. We know from our literature survey, that almost any system designed for Facial Expression Recognition has these three important steps:

- Image pre-processing
- Feature Extraction
- Expression Classification

On similar lines, a system has been proposed using these steps as shown in Figure 1.

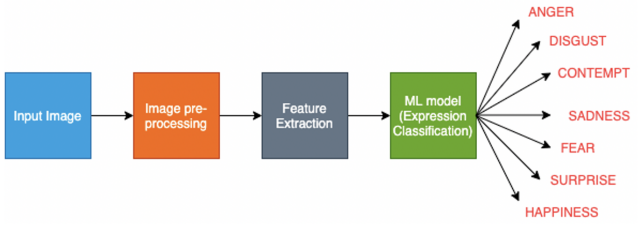

Figure 1: Block diagram of the proposed FER methodology

The way in which image is pre-processed and how the features are extracted makes it a novel technique. Human facial expressions are represented using facial landmark based feature vector. And, Support Vector Classification technique is used to recognise the facial emotions. Python programming language has been used in order to develop the entire system with extensive usage of OpenCV for performing image manipulation tasks. Predominant Scikit-learn python library and Dlib [11], an open source general purpose machine learning library have been used for performing the machine learning tasks. The primary usage of Scikit-learn library is to implement the Support Vector Classifier. Whereas, face detection and facial landmark points prediction is achieved using Dlib library.

3.1 Facial Expression Image Database

Support Vector Classifier (SVC) needs to be trained with a well-known facial expression database, thus, the CK+ dataset is selected. It consists of 7 emotions expressed by 210 adults viz., happiness, surprise, fear, sadness, contempt, disgust and anger. Each image is of size 48*48 pixels. Figure 2 shows a sample of images depicting all seven emotions. Participants are aged 18 to 50 years, 69% are female, 81% are Euro-American, 13% are Afro-American and 6% are other groups [12].

Figure 2: Seven expressions of CK+ database

3.2 Image pre-processing

We know that the dataset consists of images captured in wide range of lighting conditions. Thus, to ensure that all images are equalized to similar lighting conditions, Contrast Limited Adaptive Histogram Equalization (CLAHE) is performed [13] on all the images in the dataset using OpenCV built-in function. The advantage of using CLAHE compared to a general Histogram Equalization is that it doesn't consider the global contrast of the image. In case if the image is in RGB, which is not the case for this dataset, it should be first converted to grayscale and later CLAHE should be applied on it. For visualizing the methodology, input image shown in Figure 3 is considered which is in grayscale. And, Figure 4 shows the image obtained after applying CLAHE on the input image.

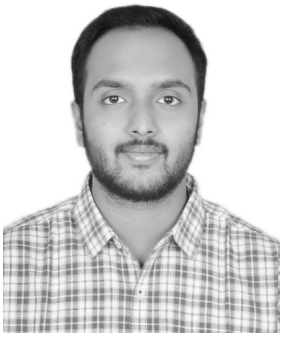

Figure 3: Input image in grayscale

Figure 4: CLAHE applied on grayscale image

3.3 Facial Landmark Prediction

Firstly, to predict the facial landmarks, the face detection algorithm must be run on the image. This was done with built-in Dlib function, which returns an object detector that is capable of identifying faces in the image. Using a classic HOG, the afore-mentioned object detector is developed. Figure 5 shows the detected face with blue bounding box.

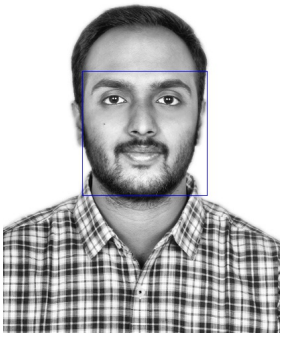

Figure 5: Detected face

Having done with the face detection, again Dlib built-in function is used to predict the facial landmark points. The 68 landmark points prediction is achieved by using the popular pretrained model available for download on the dlib website. This function internally uses the method proposed by [14] for achieving better predictions. The famous iBUG 300-W face landmark dataset is used to train this estimator as it's very robust. Figure 6 shows the predicted facial landmark points marked with red dots.

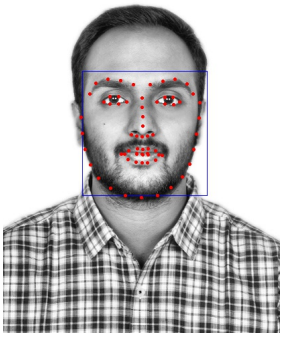

Figure 6: Predicted facial landmarks

3.4 Feature Extraction

Calculation of the feature vector that describes a person's emotion would be the most important step of facial expression identification. It's necessary to know the positions of the facial landmark points relative to each other. This is achieved by calculating mean of both the axes which results in a central point (x_{mean}, y_{mean}) as shown in Figure 7 with a large blue dot near the nose region. We can then get the position of all points relative to this central point. Now a line is drawn between the center point and each other facial landmark location as shown in Figure 7 with green lines. So, each line obtained has magnitude and direction (i.e., it's a vector) and constitutes the feature vector for both training and classification phase. The magnitude is the Euclidean distance between the points under consideration and direction is the angle made by the line with the horizontal reference axis. Therefore the feature vector can be generalized as:

feature_vector = *<point1.x , point1.y , magnitude1 , direction1 , . . . , point68.x , point68.y , magnitude68 , direction68 >*

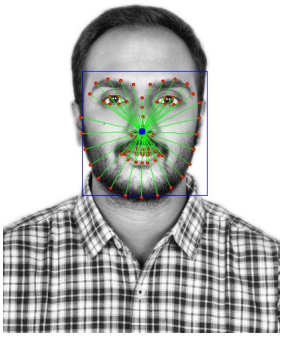

Figure 7: Feature vectors shown alongwith central point (x_{mean}, y_{mean})

3.5 Training and Classification using SVC

It is important to identify the facial expressions after constructing a feature vector. This was achieved by using Support Vector Machine (SVM) as the classification method for evaluation. SVMs are strong but versatile supervised machine learning algorithms used for classification and regression. However, they are commonly used in classification problems. Thus, SVM and SVC can be used interchangeably. SVM classifiers provide excellent accuracy and function well with high dimensional space. Basically, SVM classifiers use a subset of training points and thus use far less memory in the end. However, they have high training time which is not suitable for large data sets in practice. In our case, as we don't have a really large dataset, the training and classification can be achieved with a very good accurcay by just using a SVC. Thus, the use of Decision Trees or Random Forest is not really needed.

The SVC was implemented by using a class called "SVC" present in the "svm" module of Scikit-learn library. To provide versatility, the dataitems in each emotion of the dataset was randomly shuffled and split in the ratio of 80:20 for training data:testing data. In training phase, first each image in the training set of a particular emotion is considered and CLAHE is applied on it. Now face detection is applied to it following which the facial landmarks

can be predicted. From the obtained landmark points, the feature vector is calculated as given in the previous sub-section. Finally, SVC is trained using all the feature vectors that are calculated for the entire dataset, along with the corresponding class labels.

Each image in the testing set is taken through the same steps as that of the training phase for the calculation of the feature vector during the testing/classification phase. One important point to be noted is that training and testing set dataitems are mutually exclusive i.e., testing set consists of images which haven't been used for training. Finally, the determined feature vector for a test image is given as input to the trained SVC which predicts the emotion expressed by the test image under consideration.

4 Experimental Results Analysis

In our experiment, we had considered 80% of the images in the CK+ database for training and used the rest for testing purpose. All the seven emotions available in CK+ database was considered and the number of dataitems in each emotion is given in Figure 8. In order to keep overfitting at bay, training data was ensured to be mutually exclusive with that of the testing data.

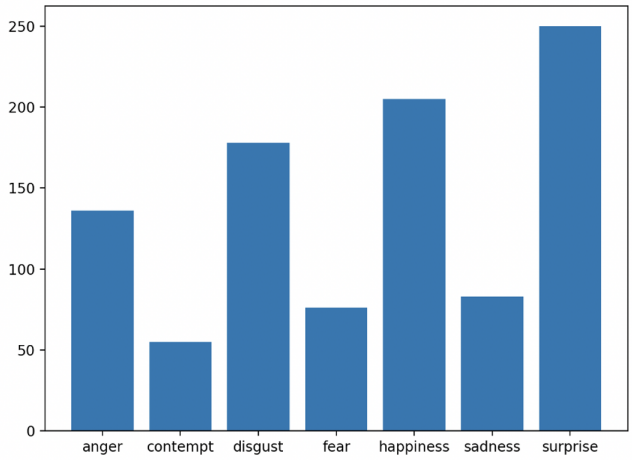

Figure 8: Distribution of the Input database

The developed model was able to classify the images in the test data into one of the 7 emotions (i.e., anger, contempt, disgust, fear, happiness, sadness, surprise) with an average success rate of 89%. This attained accuracy is much better than some of the CNN methods which generally have high training time with an extensive usage of the underlying hardware resources like CPU/GPU.

With the test samples of CK+ dataset, the proposed method was able to achieve 100% accuracy for happiness emotion as shown in the confusion matrix given in Figure 9. And the accuracy obtained for contempt, disgust, sadness and surprise emotions were also reasonably good. However, it can be observed that it's relatively hard to recognize anger and fear.

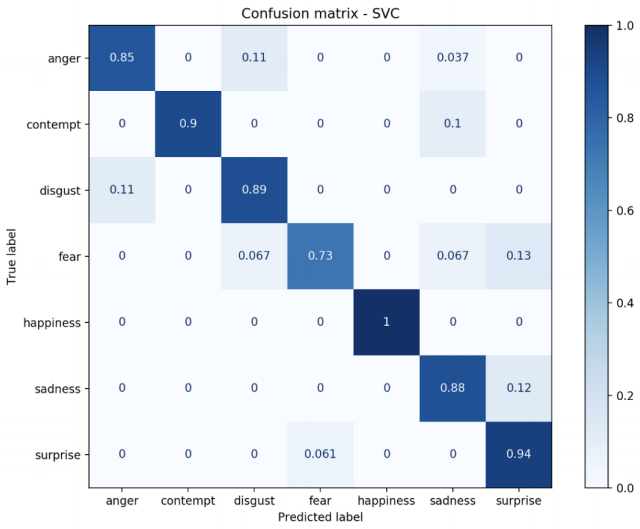

Figure 9: Confusion Matrix on CK+ dataset

To further understand why it's fairly challenging to recognize fear emotion, we need to draw some conclusions from the confusion matrix. It's observed that 13% of the images in test set of fear emotion have been misclassified as surprise. With Figure 10, we can see how challenging it is for certain images to be correctly classified even by humans. This is primarily because, the expression on the face is mixed with fear and surprise, which makes it difficult to give an accurate prediction and results in misclassification.

Figure 10: Images with true label fear incorrectly classified as surprise

5 Conclusion

The potential ability for recognizing facial expression based on facial landmarks is explored in this paper. It shows that facial expressions can be recognized by human brain using just 68 points instead of all face image pixels. Recognition of facial expressions was achieved using a Support Vector Classifier. The famous widely accepted CK+ dataset was used for both the model training and testing phase. The result on our test samples shows that the landmark based approach also has comparable performance with methods based on CNN. However, the precision of the landmark detection algorithm used, is the quintessential factor which decides the performance of the proposed method. The future work would be to explore the model for different facial poses and add support for real-time recognition.

Conflict of Interest The authors declare no conflict of interest.

Acknowledgment The authors would like to thank RV College of Engineering, Bangalore for their constant support and guidance in carrying out this project work.

References

[1] Mehrabian A., Russell J.A., An Approach to Environmental Psychology, The MIT Press: Cambridge, MA, USA, 1974.

[2] P. Ekman and W. Friesen, "Facial action coding system: A technique for the measurement of facial movement", Palo Alto, CA: Consulting Psychologists Press, 1978.

[3] Z. Zeng, M. Pantic, G. I. Roisman and T. S. Huang, "A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions," IEEE Transactions on Pattern Analysis and Machine Intelligence, **31**(1), 39–58, 2009, doi:10.1109/TPAMI.2008.52.

[4] Q. Yang, C. Li and Z. Li, "Application of FTGSVM Algorithm in Expression Recognition of Fatigue Driving," Journal of Multimedia , **9**(4), 527–533, 2014, doi:10.4304/jmm.9.4.527-533.

[5] A. R. Daros, K. K. Zakzanis and A. C. Ruocco, "Facial emotion recognition in borderline personality disorder," Psychological Medicine, **43**(9), 1953–1963, 2013, doi:10.1017/S0033291712002607.

[6] M. Wang, Z. Wang, S. Zhang, J. Luan and Z. Jiao, "Face Expression Recognition Based on Deep Convolution Network," in 2018 11th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), 1–9, 2018, doi:10.1109/CISP-BMEI.2018.8633014.

[7] L. Ivanovsky, V. Khryashchev, A. Lebedev and I. Kosterin, "Facial expression recognition algorithm based on deep convolution neural network," in 2017 21st Conference of Open Innovations Association (FRUCT), 141–147, 2017, doi:10.23919/FRUCT.2017.8250176.

[8] N. P. Gopalan, S. Bellamkonda and V. Saran Chaitanya, "Facial Expression Recognition Using Geometric Landmark Points and Convolutional Neural Networks," in 2018 International Conference on Inventive Research in Computing Applications (ICIRCA), 1149–1153, 2018, doi:10.1109/ICIRCA.2018.8597226.

[9] Sara Zhalehpour, Zahid Akhtar and Cigdem Eroglu Erdem, "Multimodal emotion recognition based on peak frame selection from video", SIViP, **10**, 827–834, 2016, doi:10.1007/s11760-015-0822-0.

[10] Hao Tang and Thomas S. Huang, "3D Facial Expression Recognition Based on Properties of Line Segments Connecting Facial Feature Points", in 2008 8th IEEE International Conference on Automatic Face & Gesture Recognition, 1–6, 2008, doi:10.1109/AFGR.2008.4813304.

[11] D. E. King, "Dlib-ml: A Machine Learning Toolkit," Journal of Machine Learning Research , **10**, 1755–1758, 2009.

[12] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar and I. Matthews, "The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression," in 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops, 94–101, 2010, doi:10.1109/CVPRW.2010.5543262.

[13] G. Yadav, S. Maheshwari and A. Agarwal, "Contrast limited adaptive histogram equalization based enhancement for real time video system," in 2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI), 2392–2397, 2014, doi:10.1109/ICACCI.2014.6968381.

[14] V. Kazemi and J. Sullivan, "One millisecond face alignment with an ensemble of regression trees," in 2014 IEEE Conference on Computer Vision and Pattern Recognition, 1867–1874, 2014, doi:10.1109/CVPR.2014.241.