

Real Time Emotion Recognizer And Classifier For Facial Expressions Based On Machine Learning approach

K.Balasaranya¹, Pacha Shobha Rani², P.Abhigna Sriya³, S.K. Rumana Taj⁴, N.Anureethika⁵

¹Assistant Professor, Department of CSE, R.M.D. Engineering College

²Associate Professor, Department of CSE, R.M.D. Engineering College

³Programmer Analyst Trainee, Cognizant Technology Solutions India Pvt. Ltd.

⁴Student, Department of CSE, R.M.D. Engineering College

⁵Student, Department of CSE, R.M.D. Engineering College

¹balasaranya1701@gmail.com,² psr.cse@rmd.ac.in, ³abhignapalagati@gmail.com,⁴rumanatajshaik@gmail.com, ⁵anureethika7091@gmail.com

Article History Received: 10 January 2021; Revised: 12 February 2021; Accepted: 27 March 2021; Published online: 28 April 2021

Abstract--There are various fields in deep learning that are known for their big success one such technique is computer vision. In computer vision the detection of face emotion is done using the convolutional neural networks technique where the model can be trained to analyze images or videos. In this paper we propose a system that recognizes the emotions that a person is having at any point of time by checking the emotion from frame to frame in a video medium. It consist of a multistage image processing to extract feature representations. We characterize six emotions based on the boundary of the face and compare it with the existing databases to get the most accurate emotion from the six.

Keywords- computer vision, face detection, frame separation, feature extraction, emotion recognizer, CNN algorithm, HAAR cascade technique.

1. Introduction

In face eyes is the most communicative and expressive part of the human being. The face can able to pass on many emotions without saying a word. Facial emotion recognition discover emotion from face image, it is a materialization of the personality and activity of a human. In the 20th century, the American psychologist Ekman and Friesen identified six classic emotions like anger, smile, fear, disgust, sadness, surprise and happy. These are the cultures of facial emotion recognition. According to research emotion in real life and education plays major role. Presently a teacher use examination, questioning, observing as sconces of feedback but these basic method mostly come with low efficiency. But these facial emotions of student the teacher can come the solution of methodology. Our system includes phases: preprocessing, face detection, feature extraction, classification and we have seven emotions: neutral, anger, happy, sadness, surprise, fear, disgust. In this project, we propose Facial Emotion Recognition using mul tipl em feature fusion. To recognize the facial expression develop the frame work which can effectively tackle the facial emotions and then through serial port the video f that emotions can be noticed.

2. Related work

According to user's review, blogs, and discussions about the face emotion recognizer. Feature extraction is an integral part of face detection which is used to find the emotion of a person. Many researchers are interested in improving the learning environment with FER.

Tang et al. [3] proposed a system which is able to analyze students' facial expressions in order to evaluate classroom teaching effect. The system is composed of five phases: data acquisition, face detection, face recognition, facial expression recognition and post-processing. The approach uses K-nearest neighbor (KNN) for classification and Uniform Local Gabor Binary Pattern Histogram Sequence (ULGBPHS) for pattern analysis.

Savva et al. [4] proposed a web application that performs an analysis of students' emotion who participating in active face-to-face classroom instruction. The application uses webcams that are installed in classrooms to collect live recordings, then they applied machine learning algorithms on its.

In [5] Whitehill et al. proposed an approach that recognizes engagement from students' facial expressions. The approach uses Gabor features and SVM algorithm to identify engagement as students interacted with cognitive skills training software. The authors obtained labels from videos annotated by human judges.

Then, the authors in [6] used computer vision and machine learning techniques to identify the affect of students

in a school computer laboratory, where the students were interacting with an educational game aimed to explain fundamental concepts of classical mechanics.

In [7] the authors proposed a system that identifies and monitors student's emotion and gives feedback in real-time in order to improve the e-learning environment for a greater content delivery. The system uses moving pattern of eyes and head to deduce relevant information to understand students' mood in an e-learning environment.

Ayvaz et al. [8] developed a Facial Emotion Recognition System (FERS), which recognizes the emotional states and motivation of students in videoconference type e-learning. The system uses 4 machine learning algorithms (SVM, KNN, Random Forest and Classification & Regression Trees) and the best accuracy rates were obtained using KNN and SVM algorithms.

Kim et al. [9] proposed a system which is able of producing real-time recommendation to the teacher in order to enhance the memorability and the quality of their lecture by granting the teacher to make modification in real-time to their non-verbal behavior like body language and facial expressions.

The authors in [10] proposed a model that recognizes emotion in virtual learning environment based on facial emotion recognition with Haar Cascades method [14] to identify mouth and eyes on JAFF database in order to detect emotions.

[11] Chiou et al. used wireless sensor network technology to create an intelligent classroom management system that aids teachers to modify instruction modes rapidly to avert wasting of time.

3. Existing system

A face image is initially examined using a landmark detector and a set of interest points is fetched. These interest points are correlated with the interest points denoting an individual's neutral expression. Deviations in interest point's location (Δx and Δy) are later examined using a classifier to certainly find out what emotion is expressed. We have chosen six people from this world, who are having seven varied expressions (among which one is taken to be neutral). Images are later treated utilizing, Application Programming Interface (API) Face++, a CNN-based technique of cognitive services in the cloud. Face images are given as input to the Face++ servers, which outputs the x and y coordinates of eighty-three interest points for every face that is identified. Within each facial expression shows there are six key expressions: bliss, anguish, curiosity, despair, fury, hatred, and contempt. Next is verification of the relationship between variations in the interest points rendered by Face ++ and the categorization done by individuals of the expressions stored in the server. For every interest point, the variation within its location when there is no expression on the face, that is when the face is normal and its location in each of the balance six expressions were computed. Hence, these variations, taken as an average among the six individuals, were plotted.

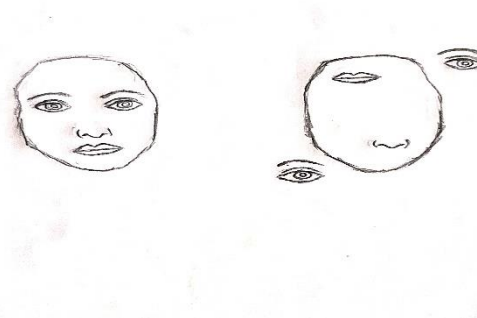


Fig 1: Image processing in existing system

In the figure 1, the boundary of both faces and the parts are the same so the CNN algorithm detects that the both pictures are same even though they actually not. So, this is the major disadvantage of the existing system that we are using only the CNN algorithm.

Here we are going to discuss about the drawbacks of the existing system and these are the following points:

1. In traditional machine learning algorithm the face emotion detection can't be captured in spontaneous manner.

2. Using PyFER the accuracy will be high but it can't detect happiness.

3. CNN do not encode the position and orientation of object.

4. Proposed system—

In this section, we describe our proposed system to analyze students' facial expressions using a Convolutional Neural Network (CNN) architecture. First, the system detects the face from the input image, and these detected faces are cropped and normalized to a size of 48×48. Then, these face images are used as input to CNN. Finally, the output is the facial expression recognition results (anger, happiness, sadness, disgust, surprise, fear, or neutral). It presents the structure of our proposed approach.

A Convolutional Neural Network (CNN) is a deep artificial neural network that can identify visual patterns from the input images with minimal pre-processing compared to other image classification algorithms. The important unit inside CNN layers is a neuron. They are connected together, in order that the output of neurons at a layer becomes the input of neurons at the next layer. In order to compute the partial derivatives of the cost function, the backpropagation algorithm is used. In fact, the CNN model has 3 modes of layers as shown in figure

1) Convolution Layer: The first layer to extract features from an input image in a proper form.

This layer maintains the spatial relationship

between pixels by learning image features using input data. It performs a dot product between two matrices, where one is the image and the other is a kernel. The convolution formula is represented in

Equation 1 :

$$\text{net}(t,f)=(x*w)[t,f]=\sum_m \sum_n x[m,n]w[t-m,f-n]$$

Where $\text{net}(t, f)$ is the output in the next layer, x is the input image, w is the filter matrix, and is the convolution operation shows how the convolution works.

2) Pooling Layer: This reduces the dimensions of each feature map but retains the most important information. Pooling is of different types: Max Pooling, Average Pooling, and Sum Pooling. The function of Pooling is to progressively reduce the spatial size of the input representation and to make the network invariant to small transformations, distortions, and translations in the input image. In our work, we took the maximum of the block as the single.

3) Fully connected layer: is a traditional Multi-Layer Perceptron that uses an activation function in the output layer. The purpose of the Fully Connected layer is to use the output of the convolutional and pooling layers for classifying the input image into various classes based on the training dataset. So the Convolution and Pooling layers act as Feature Extractors from the input image while the Fully Connected layer acts as a classifier.

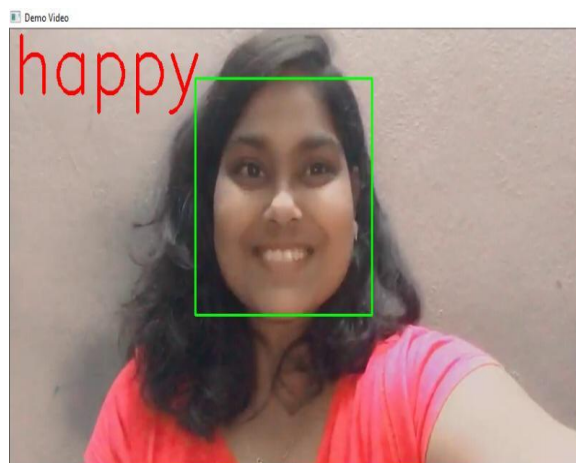


Fig 2: Result for happy emotion

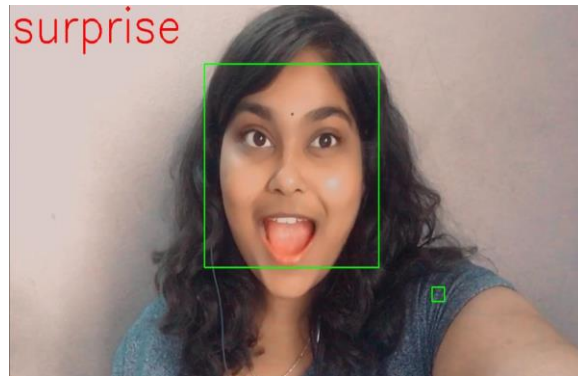


Fig 3: Result for surprise emotion

Here we are going to discuss about the advantages of Proposed system and these are the following points:

1. Because of its speed and low complexity these can be used in real time scenarios.
2. Simple classifiers are used so that the computational load need for image processing is reduced.
3. Complex image features can extracted and learnt.
4. The proposed algorithm is suitable to be implemented in embedded system with its simplicity and fast processing

5. Datasets

CK+: The CK+ dataset has 593 sequences from 123 subjects. Each sequence has 10 to 60 frames. The Dataset has people of diverse ages. Image sequences are frontal views and 30-degree views and are stored as pixel arrays of dimension 640x490 or 640x480 in gray scale or color scale. The images are labelled with seven basic emotion categories: Anger, Contempt, Disgust, Fear, Happy, Sadness and Surprise.



Fig 4: Data sets used for happy emotion



Fig 5: Data sets used for surprise emotion

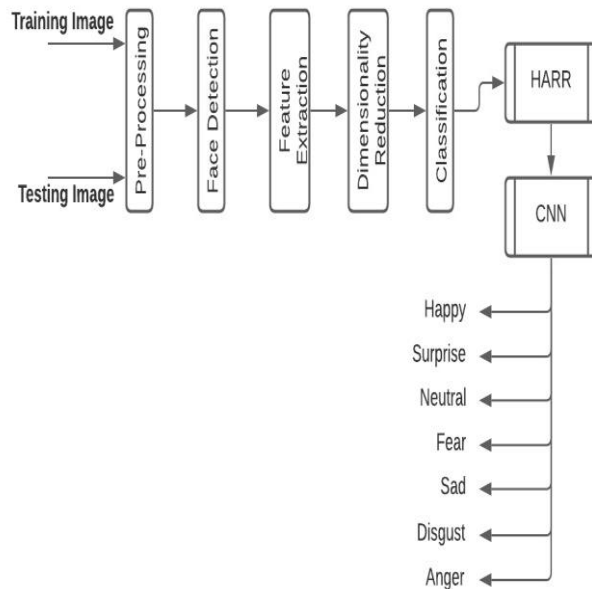


Fig 6: Block diagram for FER in real time

6. Archieture of proposed system—

PRE-PROCESSING:The aim of pre-processing is an improvement of the image data that suppresses unwilling distortions or enhances some image features important for further processing, although geometric transformations of images are classified among pre-processing methods.

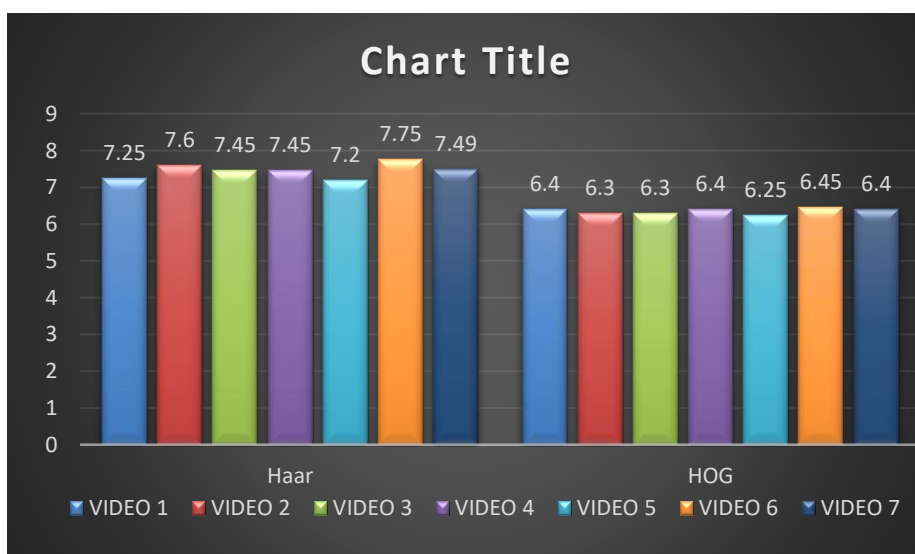
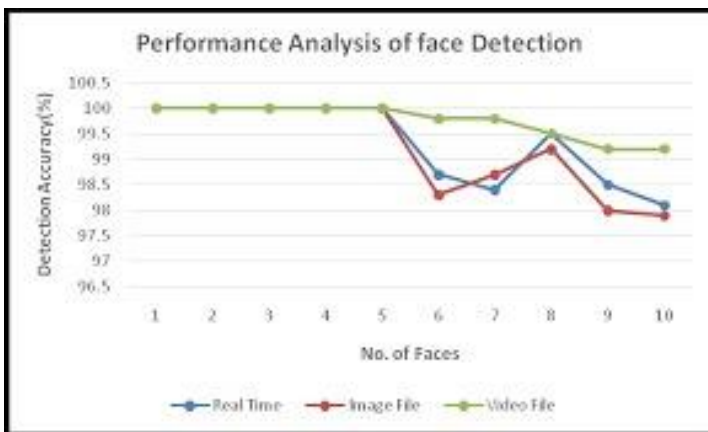
FACE DETECTION: Face detection is concerned with finding whether or not there are any faces in a given image (usually in gray scale) and, if present, return the image location and content of each face. This is the first step of any fully automatic system that analyzes the information contained in faces (e.g., identity, gender, expression, age, race and pose).

FEATURE EXTRACTION: Facial feature extraction is the process of extracting face component features like eyes, nose, mouth, etc. from human face image. In the first module, eye detector is used to detect the eye pattern using Gabor filter. In the second module, the location of eye center is found using SVM classifier to reduce the eye localization time. Finally, after locating the center and corners of eye fiducially points are found from which the face of an individual gets recognized.

DIMENSION REDUCTION: Dimensionality reduction is in the compression of image data. In this domain, digital images are stored as 2D matrices which represent the brightness of each pixel. In this paper, dimensionality reduction techniques are used to reduce the matrix representation of an image.

CLASSIFICATION: The final stage of the pipeline uses extracted Facial Features to perform face recognition (determining whose face it is) or classification (determining some characteristic of the face; for example male/female, glasses/no-glasses, etc.) All recognizers/classifiers are instances of Face Recognizer. There are a couple of default implementations, but the most common is the Annotator Face Recognizer which can use any form of Incremental Annotator to perform the actual classification.

7. Performance comparison



ALGORITHM	VIDEO 1	VIDEO 2	VIDEO 3	VIDEO 4	VIDEO 5	VIDEO 6	VIDEO 7
Haar	7.25	7.6	7.45	7.45	7.2	7.75	7.49
HOG	6.4	6.3	6.3	6.4	6.25	6.45	6.4

8. Conclusion

Through this work an inclusive practical study of facial expression recognition based on HARR features is presented. It primarily comprises of three parts, i.e. face representation, feature extraction, and classification. Face representation shows how to mock-up the face and resolves the succeeding algorithms of recognition and detection. The distinctive features of face images are extracted in feature extraction step. In the categorization or classification, the face image is evaluated among the images present in database. Face recognition gives approximately 90% of exactness; on the other hand face expression recognition also lies in the similar range of 90% accuracy. Some can build applications like:

1. Advertisement reaction model.
2. Human Computer Intelligent Interaction.
3. Emotional Templates in Video Game.

References

1. R. G. Harper, A. N. Wiens, and J. D. Matarazzo, *Nonverbal communication: the state of the art*. New York: Wiley, 1978.
2. P. Ekman and W. V. Friesen, "Constants across cultures in the face and emotion," *Journal of Personality and Social Psychology*, vol. 17, no 2, p. 124-129, 1971. [3] C. Tang, P. Xu, Z. Luo, G. Zhao, and T. Zou, "Automatic Facial Expression Analysis of Students in Teaching Environments," in *Biometric Recognition*, vol. 9428, J. Yang, J. Yang, Z. Sun, S. Shan, W. Zheng, et J. Feng, Éd. Cham: Springer International Publishing, 2015, p. 439-447.
3. A. Savva, V. Stylianou, K. Kyriacou, and F. Domenach, "Recognizing student facial expressions: A web application," in *2018 IEEE Global Engineering Education Conference (EDUCON)*, Tenerife, 2018, p. 1459-1462.
4. J. Whitehill, Z. Serpell, Y.-C. Lin, A. Foster, and J. R. Movellan, "The Faces of Engagement: Automatic Recognition of Student Engagement from Facial Expressions," *IEEE Transactions on Affective Computing*, vol. 5, no 1, p. 86-98, janv. 2014.
5. N. Bosch, S. D'Mello, R. Baker, J. Ocumpaugh, V. Shute, M. Ventura, L. Wang and W. Zhao, "Automatic Detection of Learning-Centered Affective States in the Wild," in *Proceedings of the 20th International Conference on Intelligent User Interfaces - IUI '15*, Atlanta, Georgia, USA, 2015, p. 379-388.
6. Krithika L.B and Lakshmi Priya GG, "Student Emotion Recognition System (SERS) for e-learning Improvement Based on Learner Concentration Metric," *Procedia Computer Science*, vol. 85, p. 767-776, 2016.
7. U. Ayvaz, H. Gürüler, and M. O. Devrim, "USE OF FACIAL EMOTION RECOGNITION IN E-LEARNING SYSTEMS," *Information Technologies and Learning Tools*, vol. 60, no 4, p. 95, sept. 2017.
8. Y. Kim, T. Soyata, and R. F. Behnagh, "Towards Emotionally Aware AI Smart Classroom: Current Issues and Directions for Engineering and Education," *IEEE Access*, vol. 6, p. 5308-5331, 2018.
9. D. Yang, A. Alsadoon, P. W. C. Prasad, A. K. Singh, and A. Elchouemi, "An Emotion Recognition Model Based on Facial Recognition in Virtual Learning Environment," *Procedia Computer Science*, vol. 125, p. 2-10, 2018.
10. C.-K. Chiou and J. C. R. Tseng, "An intelligent classroom management system based on wireless sensor networks," in *2015 8th International Conference on Ubi-Media Computing (UMEDIA)*, Colombo, Sri Lanka, 2015, p. 44-48.
11. I. J. Goodfellow et al., "Challenges in Representation Learning: A report on three machine learning contests," arXiv:1307.0414 [cs, stat], juill. 2013.
12. A. Fathallah, L. Abdi, and A. Douik, "Facial Expression Recognition via Deep Learning," in *2017 IEEE/ACS 14th International Conference on Computer Systems and Applications (AICCSA)*, Hammamet, 2017, p. 745-750.
13. P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, Kauai, HI, USA, 2001, vol. 1, p. I-511-I-518.
14. Y. Freund and R. E. Schapire, "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting," *Journal of Computer and System Sciences*, vol. 55, no 1, p. 119-139, août 1997.