International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 04 | Apr 2020 www.irjet.net p-ISSN: 2395-0072

Real Time Emotion Recognition System

Manthan Bhanushali[1], Sagar Bhanushali[2], Nimeesh Bagwe[3], Ashwini Deshmukh[4]

[1]*B.E Student, Shah and Anchor Kutchhi Engineering College, Chembur Mumbai India*
[2]*B.E Student, Shah and Anchor Kutchhi Engineering College, Chembur Mumbai India*
[3]*B.E Student, Shah and Anchor Kutchhi Engineering College, Chembur Mumbai India*
[4]*Professor, Dept. of IT Engineering, Shah and Anchor Kutchhi Engineering College, Chembur Mumbai India*

---***---

Abstract - *Facial expression recognition in the computer vision community was recognized as an important research subject. Consequently, many progress has been made in this sector. Emotions are expressed in words, hands and movements and facial expressions of the body. The extraction and understanding of emotion is therefore of great importance for the interaction between human communication and machine communication. The challenge involves face recognition, correct data representation, correct classification systems, accurate database, etc. This paper discusses the progress made in this area, as well as the different methods used to identify emotions. Thus, we recognized seven emotions such as Angry, Disgust, Fear, Happy, Sad, Surprise, and Neutral. Facial expression prediction preparation and research data sets are from FER 2013, which incorporate geometric features which appearance features. The main aim of the paper is to introduce a method of emotion detection in real time*

Key Words: **Emotion Recognition, CNN, DeepLearning, Classification, Opencv, Face Recognition, Machine Learning.**

1. INTRODUCTION

Studies has shown that more than 90% of our communication can be non-verbal, but technology has been struggling to keep up. Facial emotion recognition is the mechanism of identifying human emotions from the facial expressions. AI can identify emotions by learning what each face expression means and applying that knowledge to the new information that is provided to it[2]. Facial emotions are important factors in human communication that help us to understand other people's intentions. In general, people use facial expressions and vocal tone to infer the emotional states of other people, such as happiness, sorrow and rage. According to different surveys, verbal components express one-third of human communication, and nonverbal components express two-thirds. Interest in automatic facial emotion recognition (FER) (Expanded version of the acronym FER is different in each paper such as facial emotion recognition and facial expression recognition[1] .While different sensors such as an electromyograph (EMG), electrocardiogram (ECG), electroencephalograph (EEG), and camera can be used for FER inputs, a camera is the most promising type of sensor since it provides the most insightful clues for FER and does not need to be worn. Deep-learning-based FER[4,5] approaches significantly reduce the dependency on face-physics-based models and other pre-processing techniques by allowing "end-to-end" learning to take place directly from the input images in the pipeline. Among the various deep-learning models available, the most common network model is the convolutionary neural network (CNN)[4,5], which is a specific type of deep learning. In CNN-based approaches the input image is transformed into a feature map by means of a filter collection in the convolution layers. FER can also be divided into two groups according to whether it uses frame or video images.First, static (frame-based) FER relies solely on static facial characteristics obtained by extracting handcrafted features from selected frames of peak image sequences[5]. Second, dynamic (video-based) FER makes use of spatio-temporal features to capture the dynamics of expression in sequences of facial expression [4,5].Although dynamic FER has a higher recognition rate than static FER because it provides additional temporal knowledge, it does suffer from a few disadvantages.

2. LITERATURE SURVEY

An LBP cascade can be programmed to work similarly (or better) to the Haar cascade, but out of the box, the Haar cascade is about 3 times slower, and depending on the results, about 1- 2 percent better at reliably detecting a face's position. This improvement in accuracy is very notable considering that face detection will work within the precision range of 95 percent [4]. LBP is quicker (several times faster) but less reliable. (10-20% less than Haar).Because of the high execution time it takes, it is not very effective for real-time applications to use Haar classifiers, although CNN tends to collect information in shorter amounts of time while allowing to adapt to shifts or differences in the workspace over 300ms.[3]

Similar to CNNs, hair classifiers have the advantage of being able to perform simultaneous identification and classification, which removes the need to develop an external identification algorithm that extracts the image from the device and inserts it into a classifier where information losses that occur due to similarities between the target object and the context[4,5]. Furthermore, Haar classifiers do not need adjustments in their configuration to enhance recognition efficiency, as they do when specifying a CNN's architecture, but a variance in their parameters to adjust the number of steps and positive and negative training images, making them simpler to incorporate than a haar classifier.[5]

International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 04 | Apr 2020 www.irjet.net p-ISSN: 2395-0072

SVM and DBM, both excellent methods in general and with the goal of building the predictive framework. The highest output of Occurrence Detection of AUs is obtained by emotional facial classification method with SVM when comparing the experiment results of various prediction systems.[1]To increase prediction system output. We need to incorporate and compare more fusion methods.[1] PCA and negative matrix factorization (NMF), A classification algorithm using K-Nearest neighbour algorithm is used to evaluate suitable facial expressions.[2] Predicting facial expression more effectively and in the shortest possible time and processing [2]. The Haar-like features detect a facial area. Then, the facial region of interest (ROI) is reset. HOG features are extracted from the new facial ROI. FER is performed on the extracted HOG features based on SVM.[3] FER approaches jointly use CNNs and HOGs to generate feature vectors which are then sent for classification to an SVM.[3] The textural characteristics of the regions are derived from the Local Binary Patterns (LBP). Latest research shows how deep learning models on facial expression recognition could be implemented, iteratively executing the three training stages in a single loopy system. Burkert et al have proposed a facial expression recognition system for CNN [4] Extending the pattern to various face locations and testing the performance of pre-trained models for facial expression recognition, such as VGGNet. [4] Google's deep-learning open library "Tensor flow" for emotional recognition, by adding rigorous CNN to image recognition.[5]To increase the recognition performance by adding more emotional datasets and changing certain parts of the deep-learning algorithm and understanding the intention of the user and the emotional awareness of the current user.[5]AAM used a set of attribute classification schemes that defines the six fundamental emotions. FACS parameters for certain values to be stored.[6] Integrated device improvisation for the most usable user interface.[6]

3. PROPOSED WORK

The data collection used to introduce the FER program was the FER2013 dataset from the FER competition at Kaggle . The dataset consists of 35,887 images numbered, divided into 3,589 test images and 28709 train images. The dataset consists of a further 3589 private test files, on which the final test during the competition was carried out. The scale of the images in FER2013 dataset and are black and white images. The FER input picture may include noise, which can differ in illumination, scale, which colour. Few preprocessing operations were performed on the image to get precise and quicker tests on the algorithm.

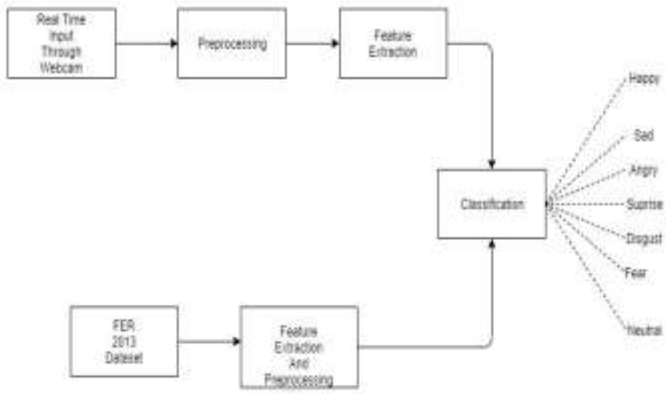

Fig 1:- Flow diagram

The primary stage for any FER program is face recognition. Haar cascades is used for face-detection (Viola & Jones, 2001). The Haar cascades, also known as the Viola Jones detectors, are classifiers that identify an entity in an picture or video they were prepared for. Haar cascades have proved to be an efficient way of identifying objects in photographs that have high precision. Hair sensors sense three dark regions on the forehead, eyebrows for example.The system classifies the image into one of the seven universal expressions - Happiness, Sadness, Anger, Surprise, Disgust, Fear, and Neutral as labelled in the FER2013 dataset. The training was done using CNN, which are a category of neural networks proved to be productive in image processing[4,5]. First, the dataset was separated into testing and evaluation datasets, and then tested on the testing collection. The retrieval of functionality on the data was not performed until it was loaded into CNN[4,5]. The method adopted was to experiment on the CNN with various architectures, to achieve improved precision with the validation package, with limited overfit.

4. EXPERIMENT

1. Dataset:-
The dataset being used in as mentioned above in proposed work i.e FER2013 dataset from Kaggel.

2. Facial Recognition process:-
The FER cycle is comprised of three stages. The preprocessing step entails assembling the data in a manner that operates on a wide-ranging algorithm and delivers accurate results. The face is identified from photographs captured in the facial recognition point, in real time. The emotional classification process consists of a CNN algorithm which classifies the image input into one of seven categories[4,5].

3. Preprocessing:-
The preprocessing techniques used are grayscale image transfer, normalization, and image resizing.
A. Normalization - The normalization of an image is performed to eliminate differences in light and to achieve an better facial picture.[3]

B. Grayscaling - Grayscaling is the method of transforming a colored image input into an image the pixel value of which depends on the light intensity of the source. Grayscaling is performed because it is difficult to process colored pictures by an algorithm.
C. Resizing - To delete redundant portions of the image the image is resized. This decreases the required memory and increases the speed of calculations.

4. Face Detection:-
The computer is trained to detect two dark regions on the face, and use fast pixel calculation to decide their location .Haar cascades effectively eliminate the unwanted background data from the image and distinguish the facial area from the image. The face detection mechanism was introduced in OpenCV using the Haar cascade classifiers.[3]

5. Emotion Classification:-
In this step,. The classification process for emotions consists of the following steps.

A. Split the Data
The data set in the FER2013 dataset was divided into three categories: Practice, PublicTest, and PrivateTest according to the label "Using". The Training and PublicTest set was used to create a configuration, and the PrivateTest set was used to evaluate the configuration.

B. Train and Generate the model
The architecture of the neural network consists of the following stratum:

i. Convolution Layer
A randomly mounted learnable filter is slid in the convolution layer, or transformed over the data. The procedure performs the dot product between the filter and each local data area.[4] The output is a multi-filter 3D volume which is also called the function diagram.

ii. Max Pooling
The pooling layer is used to reduce the input layer's spatial size to reduce the input value and the expense of the computation.

iii. Fully connected layer
Each neuron from the preceding layer is connected to the output neurons in the completely connected layer. The size of the final output layer is equal to the number of classes which classify the input image.

iv. Activation function
To minimize overfitting, the activation functions are included in. The ReLu activation function has been used in the CNN architecture. The drawback of the ReLu activation function is that its gradient is always equal to 1, meaning that the rest of the error is transferred back during back propagation

v. Softmax
The softmax function takes a vector of N real numbers and normalizes that vector into a range of values between (0, 1).

vi. Batch Normalization
The batch normalizer speeds up the exercise process and executes a transition that keeps the mean activation close to 0 and the activation standard deviation around 1.

C. Evaluating the model
The model generated during the training phase was then evaluated on the validation set, which consisted of 3589 images.

D. Using model to classify real time images
The transfer learning principle can be used to identify emotion in real-time images captured. The model developed during the training phase is pre-trained weights and values that can be used to detect a new facial expression problem. As the produced model already contains weights, FER becomes quicker for images in real time.

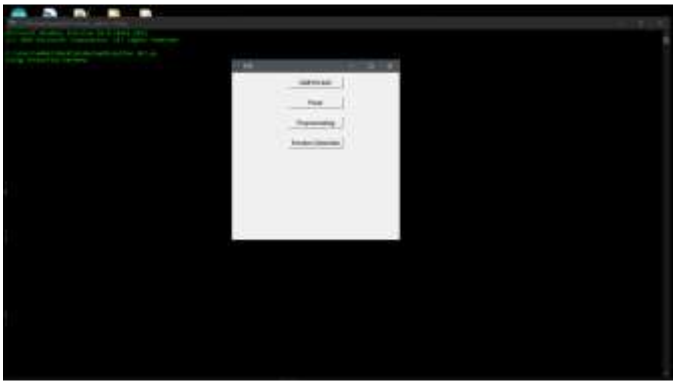

Fig 2:- Graphical User Interface

Fig 3:- Grayscale conversion

Fig 4:- Emotion detection

Results were obtained by experimentation with the algorithm CNN. The loss over training and test set has been found to decrease with each epoch. After 40 epochs the accuracy of trained model is 76. The batch size was 64, which has been held constant over all experiments. The following changes were made to achieve successful performance in the neural network architecture:

1) Number of epochs: The precision of the model has been found to increase with an increasing number of epochs. A large number of epochs however resulted in overfitting. It was concluded that eight epochs produced minimal overfitting and high precision.
2) Number of layers: The architecture of the neural network is composed of three hidden layers and one single fully connected layer. A total of six layers of convolution, were constructed using 'relu' as the activation function.
3) Filters: The accuracy of the neural network on the dataset varied according to the number of filters applied to the image. The number of filters for the network's first two layers was 64, and the network's third layers were kept at 128.

We also found that the system's accuracy improved slightly if the image quality was higher or the image was taken from a better camera, as our algorithm first detects faces and then crops the faces detected. Cropping photos decreases the quality and thus less accuracy has been expected for students who were far from the camera.

5. CONCLUSIONS

Face detection and understanding of emotions are very challenging problems.They need a strong effort to improve the face detection and emotion recognition output measurements. This field of emotion recognition is gaining interest in various domains such as gaming, software engineering, and education because of its applications.The intensity labels of RGB databases are insufficient, making it less convenient for the experiments to be conducted, and thus compromising the performance. The disadvantages of thermal databases are that they do not work with variation in pose, temperature variation, aging and different scaling (e.g. identical problem with twin). It is difficult to catch disguises if the individual has put glasses on. Thermal images have an extremely low resolution which affects the quality of the database. The 3D databases for conducting experiments and enhancing accuracy are not available in abundance. Also listed were the accuracies of different algorithms with these databases, which showed that there is room for improvement regarding accuracy in the field of emotion recognition and for detection of subtle micro-expressions.

REFERENCES

[1] Ma Xiaoxi, Lin Weisi, Huang Dongyan, Dong Minghui and Haizhou Li, “Facial Emotion Recognition” , 2017 IEEE 2nd International Conference on Signal and Image Processing.

[2] Jiequan Li and M. Oussalah, “Automatic Face Emotion Recognition System”.

[3] Sanghyuk Kim, Gwon Hwan An and Suk-Ju Kang, “Facical Expression Recognition System using Machine Learning”. ISOCC 2017 IEEE.

[4] Abirr Fathalah, Lotfi Abdi and Ali Douik, “Facial Expression Recognition via Deep Learning”, 2017 IEEE/ACS 14th International Conference on Computer Systems and Applications.

[5] Hyeon-Jung Lee and Kwang-Seok Hong, “A Study on Emotion Recognition Method and Its Application Using Face Image”, ICTC 2017 IEEE.

[6] Ashwini Ann Varghese , Jacob P Cherian and Dr. Jubilant J Kizhakkethottam, “Overview on Emotion Recognition System”, 2015 International Conference on Soft-Computing and Network Security (ICSNS -2015), Feb. 25 – 27, 2015, Coimbatore, INDIA.