Facial Emotion Recognition in image sequence Using Machine Learning Algorithms

[1]K.Swarupa Rani, [2]D.Leela Dharani, [3]P.Ravi Prakash,[4] K.Govinda Raju

Assistant Professor
[1]IT Dept,
[1]PVPSIT,[4] MVR College of Engineering, Vijayawada, India

Abstract: Facial expressions play an extremely important role in conveying human emotions. Facial expressions can be considered not only as the most natural form of displaying human emotions but also as a key non-verbal communication. As there is a greater use of human-machine interactions these days, it is also important for machines to interpret the facial expressions. Facial expression analysis is pertinent in emerging fields such as interactive games (for instance, the games played using Microsoft Kinect), online education, entertainment, autonomous driving, analysis of viewer reaction to advertisements, etc. Similarly, analysis of facial expressions of drivers would help in determining their stress level and such information could be used to alert drivers if they are stressed and in a state unsafe for driving. With these applications in mind, this paper describes our attempt to learn facial expressions from an image using machine learning algorithms. The ability to recognize different facial expressions could also improve technology that recognizes to whom specific faces belong.

Index Terms **-** Face detection, Feature extraction, Emotion classification and CNN

I. BACK GROUND AND RELATED WORK:

Many factors contribute in conveying emotions of an individual. Pose, speech, facial expressions, behavior and actions are some of them. Facial expressions are important in facilitating human communication and interactions. Facial expressions can be considered not only as the most natural form of displaying human emotions but also as a key to non-verbal communication. Also, they are used as an important tool in behavioral studies and in medical rehabilitation.

In communicating with others, humans can recognize emotions of another human with a considerable level of accuracy. The problem of automatic recognition of facial expressions is still an ongoing research. This system proposes an automatic facial expression recognition system, capable of distinguishing the six universal emotions: disgust, anger, fear, happiness, sadness and surprise. It is designed to be person independent. As there is a greater use of human-machine interactions these days, it is also important for machines to interpret the facial expressions and we would be able to attain accuracy that is virtually comparable to the human perception. The ability to recognize different facial expressions could also improve technology that recognizes to whom specific faces belong to.

In general, a person's emotions are recognized through their behavior or through some verbal communication. Humans detect and interpret faces and facial expressions in a scene with little or no effort. Still, development of an automated system that accomplishes this task is rather difficult. Recognizing one's emotions could be useful in various contexts. This may not be that effective as only one's perception is followed.

II. PROPOSED METHOD

As the machine-human interactions are emerging day by day, it would be greatly useful if the machine could predict the emotions of a person by studying their facial expressions. This is a system which automatically recognizes the emotion of a person. This includes the following works: Face detection, Facial feature extraction and Emotion classification. In this paper, we report on several advances we have made in building a system for classification of facial expressions from real time image capturing from webcam as input.This system had a wide variety of applications like in suspect detection systems etc.

Algorithm

- First, we use **haar cascade** to detect faces in each frame of the webcam feed.
- The region of image containing the face is resized to **48x48** and is passed as input to the ConvNet.
- The network outputs a list of **softmax scores** for the seven classes.
- The emotion with maximum score is displayed.

Applications:

It can be used in suspect detection systems in order to detect whether they are speaking truth or lies by studying their facial expressions. It is used to treat autism in children because this helps in studying their expressions and give a good therapy. It is used in the field of medicine i.e. in rehabilitations, counseling etc. It can also be used in e-learning in detecting the state of learner and this helps in adjusting the presentation style of tutor.

Face Detection

Given an image, detecting the presence of a human face is a complex task due to the possible variations of the face. The different sizes, angles and poses a human face might have within the image can cause this variation. The emotions which are deducible from the human face and different imaging conditions such as illumination and occlusions also affect facial appearances. In addition, the presence of spectacles, beard, hair and makeup have a considerable effect in the facial appearance. The approaches of the past few decades in face detection can be broadly classified in to four sections: knowledge-based approach, feature invariant approach, template based approach and appearance-based approach. Though various knowledge based and template based techniques can be developed for face location determination, a feature invariant approach based on the skin color is selected as the first method due to its flexibility and simplicity. When locating the face region with skin color, several algorithms can be found for different color spaces.

Feature extraction:

The face region extracted is further processed to extract the feature points required for the emotion classification stage. In this stage several feature points that were identified as important were extracted to use in the classification stage Feature extraction stage can be separated into two sections: Feature region extraction, Feature point extraction. Here sobel method is used in obtaining the edge images. These obtained edge images are then dilated and the holes are filled. The result edge images are used in refining the eyebrow regions. The face detected in the previous stage was further processed to identify eye, eyebrows and mouth regions. Finally, a corner point detection algorithm is used to obtain the required corner points from the feature regions.

a) Eye Extraction:
b) Eyebrows Extraction
c) Mouth Extraction

After feature regions are identified, these regions are further processed to extract the necessary feature points. The harris corner point detection algorithm is used to obtain the left and right most corner points of the eyes. Then the midpoint of left and right most points is obtained. This midpoint is used to obtain the top and bottom corner points. Finally after obtaining the top, bottom, right most and left most points, the centroid of the eyes is calculated. The point in the eyebrow which is directly above the eye center is obtained by processing the information of the edge image.

Emotion Classification Stage:

The extracted feature points are processed to obtain the inputs for the neural network. The neural network is trained so that the emotions neutral, happiness, sadness, anger, disgust, surprise and fear are recognized.

The inputs given the neural network are as follows.

Left eye height + Right eye height) / 2
= [(c4 - c3) + (d4 - d3)] / 2
Eye width = (Left eye width + Right eye width) / 2= [(c2- c1) + (d1 - d2)] / 2
Mouth height = (f4 - f3) , Mouth width = (f2 - f1),

Eye center to Mouth center height = [(f5 - c5) + (f5 - d5)] / 2

Left eye center to Mouth top corner length
Left eye center to Mouth bottom corner length
Right eye center to Mouth top corner length
Right eye center to Mouth bottom corner length
Eye width/ Eye height
Mouth width / Mouth height
Eyebrow to Eye center height / (Eyebrow to Eye center height + Eye center to Mouth center height).

During the evaluation phase a naive bayes classification and a manual classification was used as a benchmark evaluation for the system. Naive bayes is a simple probabilistic classifier based on the bayes theorem. The naive bayes classifier is trained using the data of the 630 training images in a supervised learning setting.

Training Model:

1. We are providing an image as an input.

2. The inputted image is then processed for

- Face detection
- Feature extraction

- Emotion classification

3. For face detection, the HAAR filter in open cv is used to automate face finding.

4. For feature extraction, corner point detection algorithm and other algorithms are used.

5. Neural networks are inputted with feature points to train the algorithms.

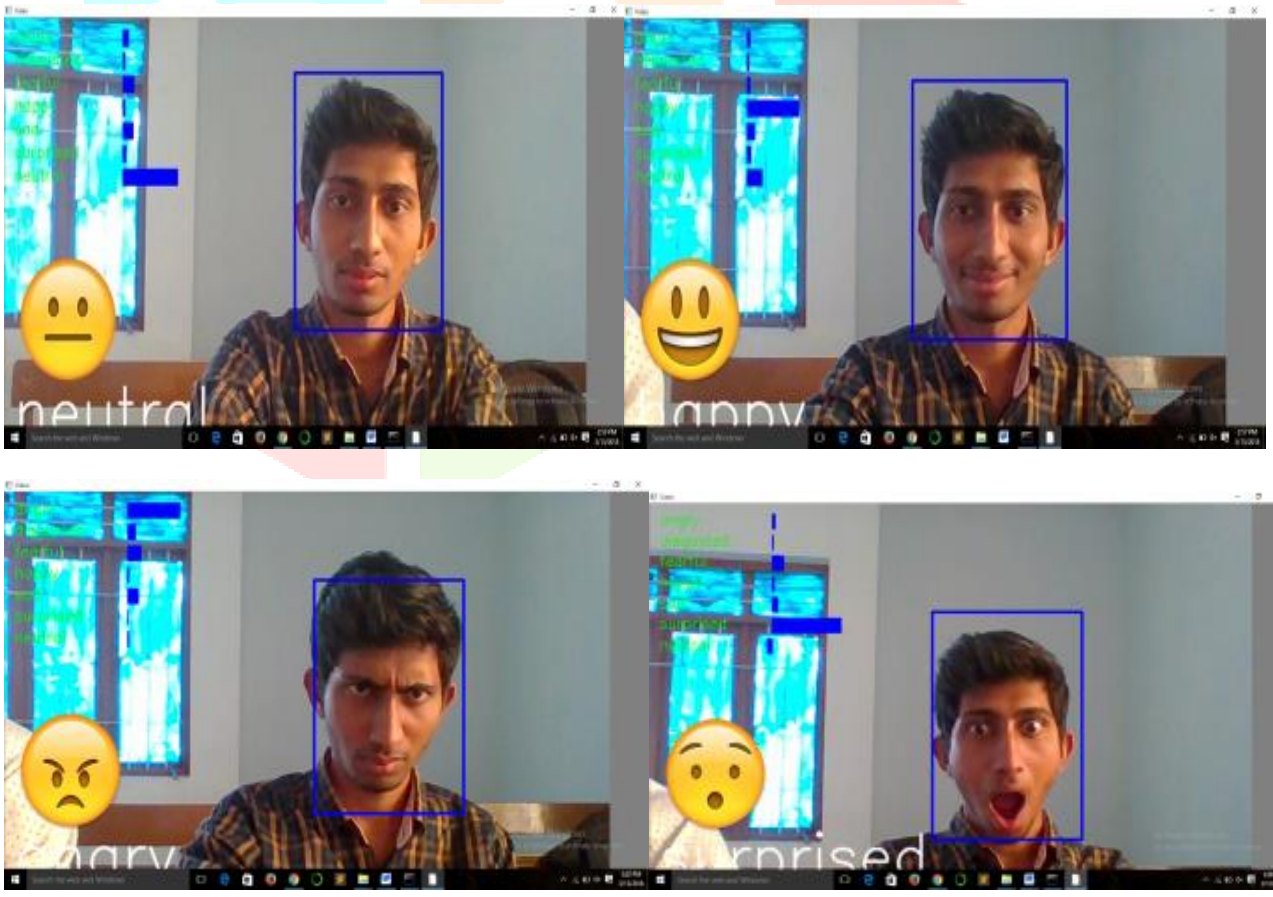
6. Classifiers are used to classify into any of the universal emotions by comparing with the trained dataset and thus to evaluate the result.

Live Application

As is already mentioned, live emotion recognition through video is one of the most important key-points in human-machine interaction. To show the capabilities of the obtained network, an application is developed that can directly process webcam footage through the final model.

With use of the afore mentioned Open CV face recognition program , the biggest appearing face from real-time image is tracked, extracted, and scaled to usable 48x48 input. This data is then fed to the input of the neural network model, which in its turn returns the values of the output layer. These values represent the likelihood that the each emotion is depicted by the user. The output with the highest value is assumed to be the current emotion of the user, and is depicted by an emotion on the left of the screen. Figures 1 to 5 shows the live application and the interaction with the authors of this research and with live capturing image from web cam.

III. EXPERIMENTAL RESULTS:



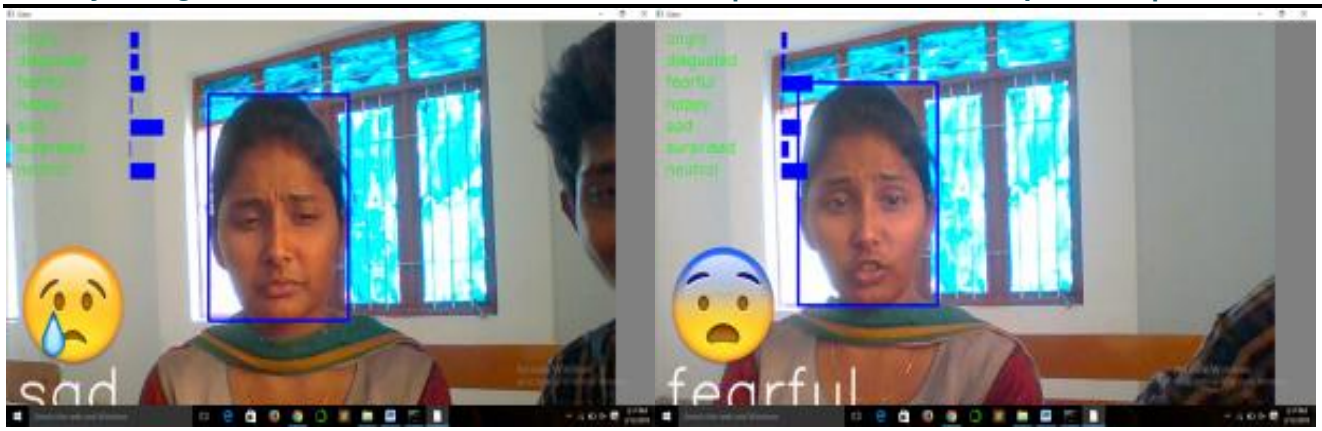
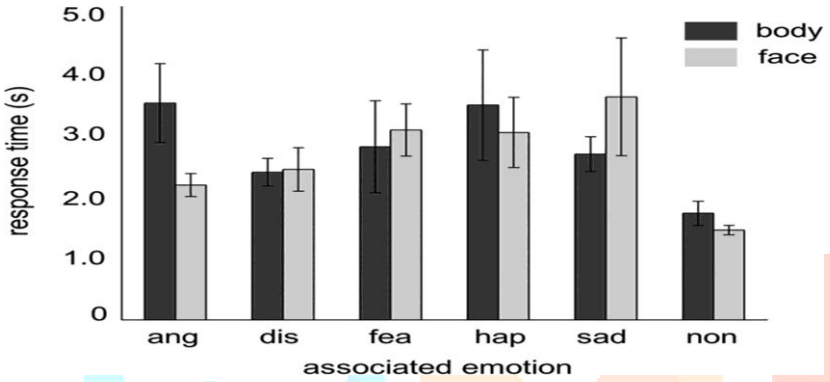
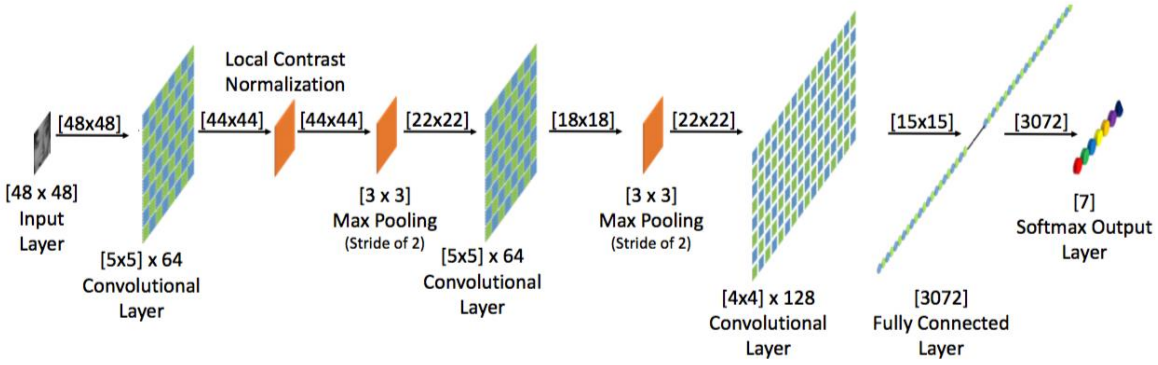


Figure 1. Sample Data Set



Convolution Neural Network (CNN):



A CNN consists of a lot of layers. These layers when used repeatedly, lead to a formation of a Deep Neural Network. Three main types of layers used to build a CNN are:

1. Input: This layer holds the raw pixel values of image.

2. **Convolutional Layer**: This layer gets the results of the neuron layer that is connected to the input regions. We define the number of filters to be used in this layer. Each filter may be a 5x5 window that slider over the input data and gets the pixel with the maximum intensity as the output.

3. Rectified Linear Unit [ReLU] Layer: This layer applies an element wise activation function on the image data. We know that a CNN uses back propagation. So in order to retain the same values of the pixels and not being changed by the back propagation, we apply the ReLU function.

4. **Pooling Layer:** This layer perform a down-sampling operation along the spatial dimensions (width, height), resulting in volume.

5. **Fully Connected Layer**: This layers is used to compute the score classes i.e which class has the maximum score corresponding to the input digits.

Dataset:

The choice of images used for training is responsible for a big part of the performance of the eventual model. This implies the need for a both high qualitative and quantitative dataset. For emotion recognition, several datasets are available for research, varying from a few hundred high resolution photos to tens of thousands smaller images. The three we will discuss are the Facial Expression Recognition Challenge (FERC-2013) [8], Extended Cohn-Kanade (CK+) [12], and Radboud Faces Database (RaFD).

IV. SCOPE FOR FURTHER DEVELOPMENT

Our system displays the emotion as based on the expression given by the user. As of now, it gives the type to emotion on the screen. We can extend this project by giving this emotion type as input to the automatic music player. By this, we can play the music as per the mood of the user. In the same way, we can use this in many of the automation systems.

REFERENCES

[1] Bhuiyan, M. A.-A., Ampornaramveth, V., Muto, S., and Ueno, H., Face detection and facial feature localization for human-machine interface. NII Journal, (5):25–39, 2003.

[2] Brunelli, R., and Poggio., T., Face recognition: features versus templates. IEEE Transactions on Pattern Analysis and Machine Intelligence, 15(10):1042–1052, 1993.

[3] Busso, C., Deng, Z., Yildirim, S., Bulut, M., Lee, C. M., Kazemzadeh, A., Lee, S., Neumann, U., and Narayanan, S., Analysis of emotion recognition using facial expressions, speech and multimodal information. In ICMI '04: Proceedings of the 6th international conference on Multimodal interfaces, pages 205–211, New York, NY, USA, 2004. ACM.

[4] Cohen, I., Garg, A., and Huang, T. S., Emotion recognition from facial expressions using multilevel hmm. In Neural Information Processing Systems, 2000.

[5] Grimm, M., Dastidar, D. G., and Kroschel, K., Recognizing emotions in spontaneous facial expressions. 2008

[6] Rituparna Halder ESL., Sushmit Sengupta ESL., Arnab Pal ESL., Sudipta Ghosh ESL., Debashish -Real Time Facial Emotion Recognition based on Image Processing and Machine Learning. (International Journal of Computer Applications (0975 – 8887) Volume 139 – No.11, April 2016)