Emojifier — Facial Emotion Recognition

Ravinder Singh and Anandesh

***Abstract*—In facial emotion recognition system the main part is the face detection and emotion recognition which works simultaneously in a systematic way, It is a complex task because of the unwanted objects, orientation, color, etc. in an image or frame of a video and this type of real-time system requires specialized hardware. Many emotion recognition systems have been developed but in this paper emotion recognition is implemented in real time system with acceptable accuracy. Basically for FER(in many papers FER is termed as Facial Emotion Recognition or Facial Expression Recognition but in this paper the expanded form of FER is Facial Emotion Recognition because generally it is abstraction and analysis of Emotions from facial components) there are three major approaches like Conventional approach, Deep Learning based approach and a hybrid Deep Learning and Conventional combined approach. This work will prove as a brief source for anyone who is interested in FER related projects.**

***Index Terms*—Emotion Recognition, Classifiers, Feature Extraction, Neural Networks, Facial Action Coding System, HAAR Cascade, CNN, OpenCV**

I. Introduction

The development in the field of Computer Vision [1] (which focuses on more human interaction with computer) has grown intensively over past two decades. In general the people express themselves using emotions (which comes under non-verbal communication) as shown in Figure(1), body actions, voice tones, etc. Humans only expresses their one third of the feelings verbally and two third of the communication is done non-verbally. In these two past decades many effective systems and technologies have been used to categorize the emotional expressions. Most of these emotional expressions are such as neutral, anger, contempt, disgust, fear, happy, sadness, surprise, etc. But most of these systems are based on Paul Ekman's categorization scheme [2] , which states only six universal emotions like anger, disgust, fear, happy, sadness, surprise. According his studies, human emotions are inferred from facial expressions by most of the people. It's very complex to figure out the actual emotion from the FE because the way we express is more natural and most of the time human combine these six universal emotions.

Fig. 1. A sample image sequence from Cohn-Kanade data set.

By using mathematical models and algorithms a machine can recognize the human emotions and some sensor device (mostly used web camera) are used to take input images or video, which can record human gestures and movements. FER uses various approaches like Conventional, Deep-Learning and hybrid approach. This project follows the Deep-Learning approach because it uses "end to end" learning process and also reduces the use of face based models. CNN (Convolutional Neural Network) [3] is a popular deep learning architecture which is used for image classification, object detection, face recognition, etc. In CNN based approach the input image is convolved into feature maps with help of filters(also called kernels). Feature maps are generated from the results of convolution layers and max-pooling or sub-sampling occurs that reduces the spatial resolution of the feature maps. Finally, the reduced feature map is used by the CNNs fully connected neural-network to classify the image and a single image is recognized with the help of softmax. Figure(2) shows the working flow of CNN.

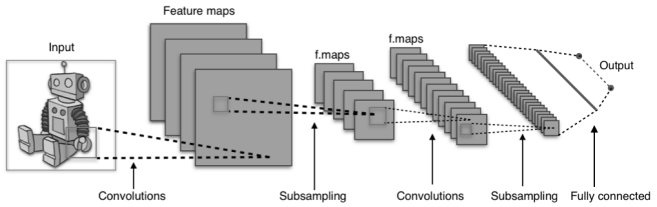

Fig. 2. Basic process that is carried out by CNN

II. Terminology

Below are some important terminologies that are used in context of FER in this project Emojifier:

1) HAAR Cascade Classifiers [4] is a classifier that is used for detecting the objects, lines, edges, etc in an image for which the classifier is trained. HAAR cascade is trained by imposing the positive images over the negative images. It requires high quality images for better accuracy and procedures are done on some specialized hardware with more GPU and CPU power.
2) Fisherface [5] method is an enhancement of the Eigenface(that uses PCA - Principal Component Analysis) method that it uses Fisher's Linear Discriminant Analysis (FLDA or LDA) for the dimension reduction.
3) FACS (Facial Action Coding System) [6] is a system that uses the facial muscles to characterize the transition in expression of a person. FACS encodes the movement of facial muscles and it is represented as Action Units(AU). With the help of FACS anyone can encode any facial movements manually.
4) AU (Action Units) & AD (Action Descriptors), [6] AU is encoded information about the rudimentary movements

Ravinder Singh (e-mail: ravinder@msi-ggsip.org)
Anandesh (e-mail: anandeshsharma@gmail.com)

of a facial muscles or group of muscles. But ADs are actually simple facial actions which involves several facial muscles (e.g. Upward rolling of eyes involves the movement of eyes).

TABLE I
THIS TABLE REPRESENTS THE ACTION UNITS(AU) FOR THE CORRESPONDING EMOTION

Emotions	Action Units (AU)
Anger	4+5+7+23
Surprise	1+2+5B+26
Disgust	9+15+16
Happy	6+12
Fear	1+2+4+5+7+20+26
Contempt	R12A+R14A
Sad	1+4+15

In Table 1, Action Units consist a group of facial muscle (e.g. For "Happy" AU is 6+12 i.e. Cheek raiser & Lip corner puller ADs are used here).

III. METHODOLOGY

In this article the system is proposed into five different stages as shown in the Figure 3.

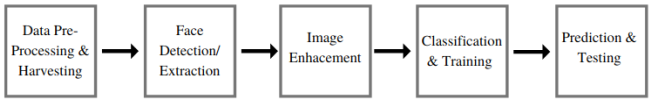

Fig. 3. Procedure followed by the system in order to recognize the facial emotions

The first task of the system is to harvest or gather data from the sources and organize it in the structured manner so that further operations can be performed on the organized data set effortlessly. The data set used by the system is CK+ (Cohn Kanade) [7] , that contains 593 video sequences on both posed and non-posed (spontaneous) emotions, 123 subjects from 18 to 30 years in age and Image resolutions of 640 × 480, and 640 × 490.

Then the second task of the system is to detect the faces with the help of cv2 and HAAR cascade(i.e. haarcascade_frontalface_default.xml) [8] and on this stage the image is normalized a little bit i.e. cropped along the face dimensions which are detected and returned by the cv2 method and converted to grayscale using cv2.cvtColor(frame, cv2.COLOR_BGR2GRAY). After, that cropped & normalized face is then saved back into the data set.

On the third stage, system will perform some image enhancements (a transparent image i.e. Emoji which represents the emotions here, will be partially drawn over the original source image) to train the classifier. Once the image preprocessing is completed, the training algorithm produces the trained prediction model. There are number of trained model that exist, but they have some strengths or weaknesses that actually depends on the situation and requirements. Some of the popular training algorithms are :

1) Multiclass SVM (Support Vector Machine) [9] is categorized under supervised learning algorithm and it analyzes and classifies the data. Also, it comes under the Convention approach for FER. This algorithm works well when classifying the human facial expressions and doesn't works well when classifying expression in uncontrolled settings.
2) RNN (Recurrent Neural Network) [10] is a type of neural network that uses the knowledge from the previous frame prediction and then predicts the current frame emotion.
3) Fisher Face [11] is one of the popular training algorithms and considered as best compared to others. The basic methodology behind Fisher Face is to perform image recognition by reducing the space dimensions of the image with the help of PCA (Principal Component Analysis) method (used in Eigen Faces) and after that applying the FDL (Fisher's Linear Discriminant) also called LDA (Linear Discriminant Analysis) method to extract features and characteristics from an image. The system we built is also using Fisher Face algorithm for recognizing the emotions.

At the fourth stage, system will generate the emotion recognition model with the help of the trained classifier and also prints the accuracy results.

Finally, system can use the newly built model to predict the emotion in real-time.

A. Modules Used

1) **Glob** is a module in python that helps in finding the path names, peek into directories, and perform task on the files. It follows the Unix shell pattern matching for the directories. Glob is used by the system to organize the data set and perpetrate necessary tasks on files.
2) **NumPy** is a general-purpose array-processing package. It provides a high-performance multidimensional array object, and tools for working with these arrays. NumPy is used by the system to work with image arrays and ndarrays.
3) **CV2** is an interface for python that is officially released by OpenCV and its basically a bundle of various methods and classes that accelerates the machine perception. CV2 is used by the system for reading the video frames, face detection using HAAR cascade classifier, emotion recognition using Fisher Face classifier, building emotion recognition model.
4) **PIL** (Pillow) is python imaging library that is used for image manipulation and enhancement. PIL is used by the system to normalize the images for training and prediction purposes.

IV. CONCLUSION

Many researches have been published over past decades on Facial Emotion Recognition in order to achieve higher accuracy, less resource requirement and robustness. This paper presented the brief overview of the approaches, algorithms, technologies used by the system we built. We described the

modules, classifiers, approaches and algorithms that are used to build this system. Also, this paper describes the data set which our system is using for training the classifiers and building the emotion prediction models. On the first test, using these 593 video sequences and Fisher Face algorithms our system gave us the accuracy of **60.3%** which clearly means that the prediction of our classifier is **39.7%** wrong. But there are still chances of improvement, we manually corrected some the samples by removing conflicting images and adding new images corresponding to the emotions. And the second time we again re-train the classifier on this improved data set, the accuracy incremented by **13.2%** i.e **73.5%**. The main objective of this paper is to describe the results of the FER approach and algorithm that we've implemented on our system.

REFERENCES

[1] G. Bradski and A. Kaehler, "Learning OpenCV: Computer vision with the OpenCV library," and others, Ed.

[2] P. Ekman, "Handbook of cognition and emotion." Wiley Online Library, 1999.

[3] S. Lawrence, C. L. Giles, A. C. Tsoi, and A. D. Back, "Face recognition: A convolutional neural-network approach," and others, Ed.

[4] "OpenCV: Face Detection using Haar Cascades."

[5] "Fisherfaces - Scholarpedia."

[6] "Facial Action Coding System - Wikipedia."

[7] 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, 13-18 June 2010, San Francisco, CA, USA*, and others, Ed. IEEE, June 2010. [Online]. Available: https://ieeexplore.ieee.org/document/5543262

[8] "Haarcascade - GitHub."

[9] "Multiclass SVMs - Stanford NLP Group."

[10] A. Sherstinsky, "Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) Network," no. 9 Aug 2018, 2018. [Online]. Available: https://arxiv.org/abs/1808.03314

[11] J. P. Hespanha, D. J. Kriegman, and P. N. Belhumeur, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection."

Ravinder Singh Asstt. Prof. & TPO at MSI, New Delhi. Received B.Sc in Electronics degree from MDU Rohtak, Masters in Electronics from Kurkshetra University and M.Phil from CDLU. Presented many papers in Singapore and Malasiya in various international conferences

Anandesh is a Linux lover, a docker & ML enthusiast, a security pentester, a problem solver, a tech speaker and a people's guy.

Currently he's pursuing Bachelors of Computer Applications (BCA) 2016-2019. And also he's affiliated to Microsoft Student Partner program.