Audience Feedback Analysis using Emotion Recognition

Mrs. Madhura Prakash
Assistant Professor,
Dept. of ISE BNM Institute of Technology
Bangalore, India

Aman Kumar, Anmol Saxena, Somil Agrawal, Twinkle Singh
Students, Dept. of ISE
BNM Institute of Technology Bangalore, India

Abstract—Image Processing is the major rising technique in the modern world of AI based systems. Emotion recognition is one of the topic inside the Artificial intelligence machine learning Techniques. Emotion Recognition is basically reading an individual's mind and present mental state by analysing the facial Expression, body posture and gesture and speech. By doing so, one can able to judge whether the individual is interested in the ongoing activity or not. The core functionality of the Audience Feedback Analysis Using Emotion Recognition is to capture and analyze the facial expression of every individual sitting in any seminar/talk/lecture/keynote and provide an an analysis of the average emotional state of the gathering during or at the end. The primary objective of the proposed system is to provide the organizer or the administrator of the event with the actual and unbiased feedback by analysing the same through the expressions of the gathering for the whole time or at some interval of time.

Index Terms—Convolutional Neural Network, Deep Neural Network, Facial Expression, Frame Extraction.

I. INTRODUCTION

The facial perception where related to the chronicity, illness and social competence, Emotions take an essential part in day- to-day life, People can recognize someone else's feelings and respond in a hasty-manner with certain circumstances. For example, "A judgment of a man using psychological study", Facial emotion recognition is a challenge because of its hazy, where features are effective for the task of which extracting effective emotional features is an open query. It is ordinarily utilized for security systems, mobile application unlocking systems as well as iris scan unlocking systems for high-tech security of latest tech for example, unique mark or eye iris recognition systems, incompletely in light of the fact the machines don't comprehend the feeling states. We can also judge a man if he is convinced for the inspirational speech or not. Emotion is a conscious experience characterized by extreme mental movement and a certain degree of pleasure or disappointment. Scientific conversations have had different implications and there is no general agreement on the defini- tion.

The facial expression is a visible manifestation of emotional state, a cognitive activity, an intention, and represents the personality and psychopathology of a person. It also has a communicative role in interpersonal relations. Facial expres- sions and gestures are included in non-verbal communication. These clues complete the speech, helping the listener to understand the meaning of the words uttered by the speaker. Mehrabian said that facial expressions have a substantial effect on the listener. The body language of the speaker is 55 percent from the message, 38 percent is transmitted by voice (such as tone, intonation and volume), and 7 percent represents the actual content of the message.

The face of a deliberate feeling is to show a psychological picture of this feeling of past or speculation, which is once again linked to a satisfactory state of joy or dissatisfaction. In-case there is a chance for that device collaborating, on the other hand, the machine learning using AI is effective in doing the job and there is a discussion in investigating human behaviour and working in a similar manner. To do this many imaginable ways we can be visualized using facial discovery to get a precise and accurate way to examine human feeling and behavior which the famous company named Volvo developed a car that conducts job interviews in Volvo's AI stunt.

Also, Nowadays getting true and unbiased response by the people or gathering for any event organized is very rare. People pretend to the organizer something but inside the true feeling is different and is hidden to everyone else outside. Getting the knowledge of this true feeling is very important for the organizer to judge his/her performance or the performances made by his/her teammates. Without this it becomes very difficult for the organizer to predict the updations or changes that are required for his actions to be better for the next time. Keeping in mind all the challenges, this paper presents a software application for Audience feedback analysis based on Emotional state of the gathering of people in any keynote/seminar/talk etc., and is intended for the organizers to get the unbiased and realistic review of the event. This application has the ability to recognize emotions from user captured images live during a live or ongoing seminar or from any uploaded video stream. Depending on the facial features of the person in the picture, the application can describe its emotional state and generate an average analysis at the end of the event or even during the event.

II. METHODOLOGY

The complete process of generating the average feedback of the audience is divided into 2 different modules 1) extracting out frames from the live streaming or from any video stream and cropping out the faces of all the individuals in those frames 2) analysing the emotion of each of the individuals separately from each frame and calculating the average emotion during the event.

A. Facial detection from input

There have been many different approaches as methods, techniques and algorithms for finding optimal, reliable, ef- ficient and effective solutions for detecting and interpreting human faces in digital images. In 1960, after a cross-cultural research, Paul Ekman proposed the idea that facial expressions of emotion are not culturally determined, but universal and defined five basic emotions: Anger, Happy, Neutral, Sad and Surprise. He and Wallace V. Friesen adopted in 1978 a system, called FACS (Facial Action Coding System), for classifying physical expressions of emotions, originally devel- oped by a Swedish anatomist named Carl-Herman Hjortsjo¨. They published an important update in 2002. Movements of individual facial muscles are encoded and they proved to be useful to psychologists and to animators.

Using FACS, any anatomically possible facial expression can be coded, splitting it into specific Action Units (AU) and their temporal segments that produced the expression. Many other researchers have used images and video processing to automatically track facial features and then to classify the various expressions. In general, the approaches are similar and search facial features using motion models of the image (op- tical flow, DCT coefficients, etc..). Based on these results the classifier is trained. The difference between existing methods is the set of features extracted from images and the classifier used (based on Bayesian or hidden Markov models). In 1990s, the research community for computer animation was faced with similar problems as in the days of pre-FACS that there is no single standard system for animating facial expressions. Pandzic and Forhheimer noted that efforts were concentrated on facial movement instead on facial parameters. Because of this Motion Pictures Expert Group (MPEG) introduced the facial animation (FA) MPEG-4 standard specifications for a standardized parametrization of facial control.

The very first step for cropping the faces of the individuals in the gathering is to extract out the frames from the input feed. Without extracting the frames, getting the cropped faces of the people is a very difficult task. As the video also is a collection of frames shown at certain rate per second, extracting out frames is precisely a simpler task. The frames can be extracted by writing a simple frame extraction program in any of the programming language. A question that can arise is that how many frames will be needed for the process. Talking precisely, this is the decision of the organizer that of how much time he/she wants the feedback of the audience. Frames extraction can be done for the entire session of the event or for any particular interval of time as well. When the next frame is to be taken after a frame can also be set accordingly. In general practice, it is strongly recommended that the frames are extracted every second so as to get the maximum accuracy for detecting the emotional status of the gathering. After the frames are extracted, the next step is to detect the faces in those frames and crop them out for analysis.

In this paper, the method used for detecting the faces from the input provided is the Deep Neural Network (DNN) algorithm of the Machine Learning practices. The choice came upon DNN algorithm because of many of the advantages it had over an other learning algorithm. Firstly, the accuracy of detecting the correct portion of the face was best when we used the DNN algorithm as compared to other techniques and algorithms such as Haar cascade technique, hbase technique etc. Secondly, the number of faces detected by the Deep Neural Network algorithm was the best as compared to all other techniques as at once this algorithm was able to detect almost all of the faces in the frame. At last, the DNN algorithm, apart from other algorithms, was able to detect the faces of the people who were not also looking towards the camera or in the direction of it. in other words, people who were looking aside or somewhere else also were able to be cropped by the DNN algorithm which was not possible using other techniques. These cropped faces of all the peoples repeatedly from all the frames are then stored in a separate folder and are used as input for detecting the emotion of each individual separately.

B. Analysing the Emotion

To understand the emotions and feelings of others, a first step is to analyse facial expressions. This is not simple for a ”machine”, which should be trained and taught a recognition technique. Any Computer Vision application starts with a preprocessing step, applied to the input image. Preprocessing involves removal conditions of light and noise, thresholding. Then the second step follows, the segmentation, extracting regions of interest from an image and isolating them to find objects of interest. For example, in a face detection system, we need to separate the faces parts from the rest of the scene. After obtaining the objects within the image, we will continue with the next step, extracting all the features of each detected object, such as surface, contour, model, etc. These features ale called descriptors and are used for training.

The proposed method was implemented using Open-CV, an open source Computer Vision and machine learning software library. Open-CV become standard in this field, one of its main advantages being that it is highly optimized and available on almost all platforms.

The Analysing algorithm behind Open-CV for the analysis of the emotion is the Convolutional Neural Network (CNN) Algorithm.

III. STRUCTURAL ARCHITECTURE

The figure 1 depicts the Structural Architecture of the project.

The step wise procedure of the method proposed is as follows:

1) The recorded video file or the Live stream of the video is inputted to the program.

2) From the input video depending upon the administrator’s need, the frames are extracted using the frame extraction

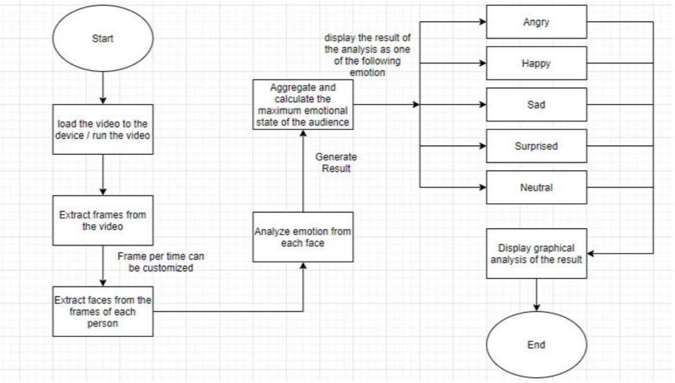

Start
load the video to the device / run the video
Extract frames from the video
Frame per time can be customized
Extract faces from the frames of each person
Analyze emotion from each face
Generate Result
Aggregate and calculate the maximum emotional state of the audience
display the result of the analysis as one of the following emotion
Angry
Happy
Sad
Surprised
Neutral
Display graphical analysis of the result
End

Fig. 1. Structural Architecture of Audience Feedback Analysis.

program. The time duration between each frame can be set by the administrator. Also the administrator has the option to limit the frames to a particular number as needed for analysis and evaluation.

3) From each frame extracted, the faces of each and every individual are cropped or extracted. It does not depend whether a person face is cropped from a frame, it will be cropped every time it occurs in any of the frame. This way every person's emotion at all the instances can be captured for the better analysis of the feedback. These cropped faces are then transformed to gray scale pictures for better analysis.

4) For each of the faces cropped the emotion is classified and is stored at a separate location of evaluation purpose.

5) The count of each emotion is also to get the maximum time of the emotion that the gathering had during the event. This calculated emotion count is then displayed to the admin- istrator in the form of bar graph showing the number of times of each emotion analyzed and also in the form of Pie Chart for the administrator to get the average emotion of the gathering of the event.

6) Based on the Calculation the most occurred emotion or the average emotion of the gathering is displayed as the result of the Project to the Administrator.

IV. IMPLEMENTATION

A. Dataset

A data set (or dataset) is a collection of data. In the case of tabular data, a data set corresponds to one or more database tables, where every column of a table represents a particular variable, and each row corresponds to a given record of the data set in question. The data set lists values for each of the variables, such as height and weight of an object, for each member of the data set. Each value is known as a datum. Data sets can also consist of a collection of documents or files. The values may be numbers, such as real numbers or integers, for example representing a person's height in centimetres, but may also be nominal data (i.e., not consisting of numerical values), for example representing a person's ethnicity. More generally, values may be of any of the kinds described as a level of measurement. For each variable, the values are normally all of the same kind. However, there may also be missing values, which must be indicated in some way In statistics, data sets usually come from actual observations obtained by sampling a statistical population, and each row corresponds to the observations on one element of that population. Data sets may further be generated by algorithms for the purpose of testing certain kinds of software. Some modern statistical analysis software such as SPSS still present their data in the classical data set fashion. If data is missing or suspicious an imputation method may be used to complete a data set.

In this project, the dataset used are 48*48 mm sized grey scaled images for people and these images are used to train the machine to detect the emotion from a picture. The training dataset is taken in gray scale images because the faces cropped from the frames will also be grey scaled and for machine feature extraction is much more easier and accurate for grey scale images rather than coloured images. The images used are of people representing emotions. These images contain a huge range of ways in which a particular emotion can be addressed by the facial landmarks of the person. The bit-depth of these images are 8 bits and these images have a resolution of 96 dpi. For all the five emotions, around 24000+ images are be given to the machine for training purpose and around 3000+ images are used for validation of the trained model. As it is a neural network program, the model is currently being trained for 50 epoch values. This epoch value can be increased to get higher accuracy if needed. the sample of these images is shown in the figure

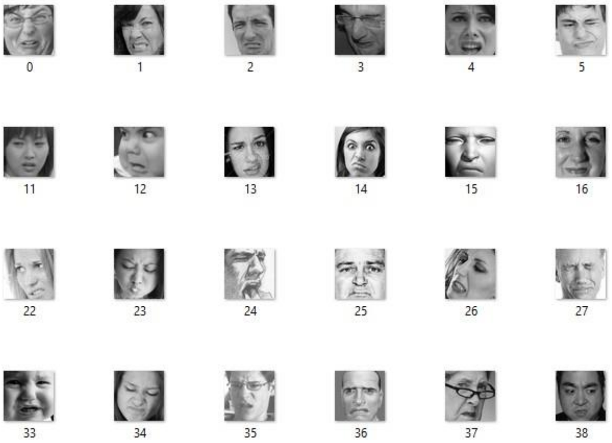


Fig. 2. Sample Dataset images.

These images represent the dataset of the emotion anger. Similarly, all the emotions have such kind of images as the dataset for the model to be trained. For each emotion there are around 4500+ images for training purpose so that the accuracy of the model can be maximum.

B. Input

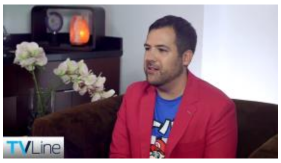
An input is something that we give to any program or any operation as the initial data that is to be processed to generate the required output. in other words we can say that, input is the feeding of information to the program to be operated upon. In this project, the input given can be given in multiple ways. The input can be either a image containing the faces of people to be analysed or it can be a video stream. As the project is based on analysing the feedback of audience in any seminar or talk, the input can be preferably a recorded video of the full event or even the live video stream of the ongoing event. The input video can be any variable size and length. The program accepts all type of video formats like mp4, mkv,
etc. of any length.

C. Frame Extraction

Generally, Video is a massive volume object with a high redundancy and insensitive information, it has complex struc- ture consists of scene, shot, frame. One of the fundamental units in the structure analysis is the Key-frame extraction, it provide us with a good video summarization and browsing a large video collections. Key-frame is a frame or set of frames that having a good representation of the entire content of small video clip. It must contain most of the salient features of the represented video clip.

Hence the first Step of the project is to generate out or
Extract out the frames from the video clip or the video stream entered as the input to the program. Frame Extraction in this project, can be considered as a separate and complete process which do not involve any involvement of any of the either algorithm used as for a feedback analysis program, every second is important to analyse the event. Thus, for frame extraction a separate set of commands are defined which allows the administrator to extract out frames at the time period he/she wants or from the full length of the video. The administrator also have the option to limit the number of frames if it is desirable.

The program was tested for a sample video inputted and
some of the frames extracted from that video are:-



These frames will be then used as an input to the emotion recognition program for the further process.

D. Faces Extraction

The next stage of the program is the cropping of the faces of each of the individual present in the frame from all the frames extracted during the last step. These cropped faces

Fig. 3. Sample Frames extracted from a inputted video.

will be stored as a image in a buffer file and will be used for analysing the emotions. Once the faces are cropped, those will be converted into grey scaled images as discussed before for the reason of better accuracy and better analysis of the emotional state. It does not depend on whether a person's face is already cropped from the previous frame or not, every time a face occurs in any of the frame will be treated as a separate image to be analysed to generate emotional status.

Some of the Sample faces extracted from the frames are:-

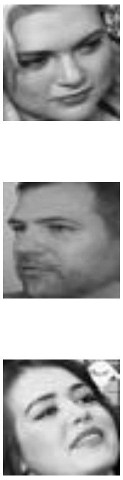


Fig. 4. Sample Faces cropped form the Extracted Frame video.

E. Emotion Generation

The final step of the program is to generate the emotional value from the images of the faces cropped. The program runs in a loop where all the faces that are extracted or cropped are analysed and emotion is generated based on the Convolutional neural network algorithm for all the images and these images are stored in a separate folder named as Emotion output which will be in the same directory where the program is stored. The images will be stored with the Emotion as the name of the image. That emotion will be with respect to the emotion the person in the image depicts.

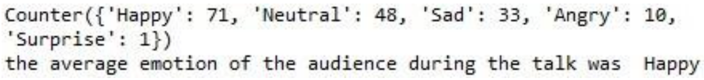
These images will have the emotion as the name of the emotions, the person in the image depicts. These emotions generated will be stored in a separate dictionary where each emotion will be counted and the number of time each emotion is occurred during the recognition process will be incremented accordingly. This way the average emotional status will be calculated as what was the maximum emotional state the gathering were in.

Fig. 5. Sample Emotion Generated images.

The Sample of all the 5 emotions analysed images are:- Similarly all the images of the faces cropped will be stored in the folder with the emotion as those image name.

V. RESULTS

After the Calculation of the emotional States, the result will show the average emotional state of the gathering and also the Graphical representation of the number of times each emotion has appeared and also the percentage of gathering having any particular emotion. The result that will be visible to the user will be these 3 content.



```
Counter({'Happy': 71, 'Neutral': 48, 'Sad': 33, 'Angry': 10,
'Surprise': 1})
the average emotion of the audience during the talk was  Happy
```

Fig. 6. Result Stating the Average Emotional State of the gathering during the event.

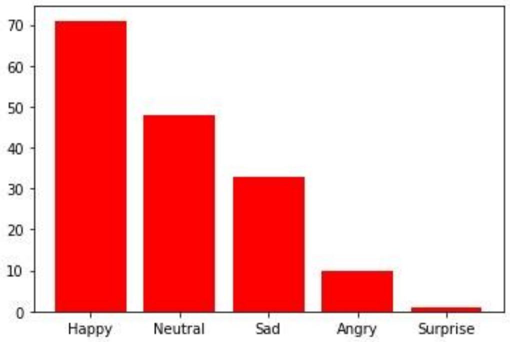


Fig. 7. Bar Graph Showing how many time a particular Emotion occurred.

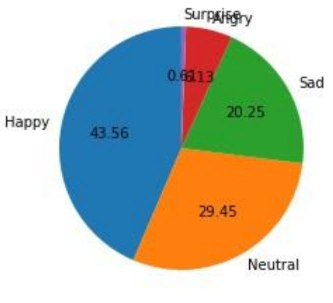


Fig. 8. Pie Chart showing percentage of gathering having particular emotion.

VI. CONCLUSION

Automatic detection of faces is difficult, practically impossible to define, existing a variety of ways in which human figures may appear in pictures. And not just because of the physiognomy traits and characteristics, but especially because of the variety in their perception in 2-D, due to position (orientation/leaning, scaling).

Automatic detection of faces in digital images is included as a feature in digital cameras, facial recognition applications included in security systems, these being just some examples most handy.

In our project, we tried to implement the full ongoing into different modules which helped us in achieving great accuracy for the purpose stated of the project.

The part where we had some difficulties was the training of the system in order to detect the emotional state. The application is used to detect emotional states based on facial features, specifically, according to face shape, position and shape of the mouth and nose. We intend to future develop the application in order to consider for emotion detection other stimuli also, such as voice and body posture and to compare our final solution with existing ones in terms of performance and emotion detection accuracy.

REFERENCES

[1] T. Shiva, T. Kavya, N. Abhinash Reddy, Shahana Bano, "Calculating The Impact Of Event Using Emotion Detection", International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-7 May, 2019.

[2] Debishree Dagar, Abir Hudait, H. K. Tripathy, M. N. Das, "Automatic Emotion Detection Model from Facial Expression", 2016 International Conference on Advanced Communication Control and Computing Tech- nologies (ICACCCT)

[3] Loredana Stanciu, Florentina Blidariu "Emotional States Recognition by Interpreting Facial Features," in The 6th IEEE International Conference on E-Health and Bioengineering - EHB 2017

[4] Binh T. Nguyen,Minh H. Trinh, Tan V. Phan and Hien D. Nguyen, "An Efficient Real-Time Emotion Detection Using Camera and Facial Landmarks,"in The 7th IEEE International Conference on Information Science and Technology Da Nang, Vietnam; April 16-19,2017.

[5] Mehmet Akif OZDEMIR, Berkay ELAGOZ, Aysegu lALAYBEYO- GLU, Reza SADIGHZADEH and Aydin AKAN, "Real Time Emotion Recognition from Facial Expressions Using CNN Architecture," 978-1-7281-2420-9/19/ 31.00 2019 IEEE.

[6] FU Zhi-Peng, ZhANG Yan-Ning, HOU Hai-Yan,"Survey of Deep Learn- ing In Face Recognition," 978-1-4799-6284-6/14/ 31.00 2014 IEEE.

[7] Williams. D. Ofor, Nuka. D. Nwiabu, Daniel Matthias,"Face Detection and Expression Recognition Using Fuzzy Rule Interpolation", Interna- tional Journal of Computer Sciences and Engineering Vol.7(5), May 2019, E-ISSN: 2347-2693.

[8] Pradeep Kumar. G. H, Ashwini. M, Divya. G. N, Manjushree. B. N,"Multiple Face Detection and Recognition in Real-Time using Open CV," International Journal of Engineering Research and Technology (IJERT) ISSN: 2278-0181, NCRTS-2015 Conference Proceedings.