

PROFESSIONAL TRAINING REPORT
at
Sathyabama Institute of Science and Technology
(Deemed to be University)

Submitted in partial fulfillment of the requirements for the award of
Bachelor of Engineering Degree in Computer Science and Engineering

By

AVIJIT SINGH
REG. NO. 38110059

KUBLAY KUMAR
REG. NO. 38110274



DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

SCHOOL OF COMPUTING

SATHYABAMA INSTITUTE OF SCIENCE AND TECHNOLOGY
JEPPIAAR NAGAR, RAJIV GANDHI SALAI,
CHENNAI – 600119, TAMILNADU

JANUARY 2022



SATHYABAMA
INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)

Accredited with Grade “A” by NAAC
(Established under Section 3 of UGC Act, 1956)
JEPPIAAR NAGAR, RAJIV GANDHI SALAI
CHENNAI– 600119
www.sathyabama.ac.in



DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

BONAFIDE CERTIFICATE

This is to certify that this Project Report is the bonafide work of **AVIJIT SINGH (Reg. No: 37110059)** and **KUBLAY KUMAR (Reg. No: 38110274)** who carried out the project entitled “**Face recognition system for sentiment analysis**” under my supervision from June 2021 to November 2021.

Internal Guide – Dr. K. Babu M. Tech., (Ph. D)

Head of the Department – Dr. R. Subhashini M.E, Ph.D.,

Submitted for Viva voce Examination held on: _____

InternalExaminer

ExternalExaminer

DECLARATION

I, **AVIJIT SINGH and KUBLAY KUMAR** hereby declare that the project report entitled “**Face recognition system for sentiment analysis**” done by me under the guidance of **Dr. K. Babu M. Tech., (Ph. D)**, at “Sathyabama Institute of Science and Technology (Deemed to be University), Jeppiaar Nagar, Rajiv Gandhi Salai, Chennai – 600119” is submitted in partial fulfillment of the requirements for the award of Bachelor of Engineering/Technology degree in Information Technology.

DATE:05/12/22

PLACE: Chennai

SIGNATURE OF THE CANDIDATE

ACKNOWLEDGEMENT

I am pleased to acknowledge my sincere thanks to **Board of Management of SATHYABAMA** for their kind encouragement in doing this project and for completing it successfully. I am grateful to them.

I convey my thanks to **Dr. T. Sasikala M.E., Ph.D, Dean, School of Computing, Dr. S. Vigneshwari, M.E., Ph.D. and Dr. L. Lakshmanan, M.E., Ph.D., Heads of the Department of Computer Science and Engineering** for providing me necessary support and details at the right time during the progressive reviews.

I would like to express my sincere and deep sense of gratitude to my Project Guide **Dr. K. Babu M. Tech., (Ph. D)** for his valuable guidance, suggestions and constant encouragement paved way for the successful completion of my project work.

I wish to express my thanks to all Teaching and Non-teaching staff members of the **Department of Computer Science and Engineering** who were helpful in many ways for the completion of the project.

ABSTRACT

This paper discusses the application of feature extraction of facial expressions with combination of neural network for the recognition of different facial emotions (happy, sad, angry, fear, surprised, neutral etc..). We use transfer learning on the fully connected layers of an existing convolution neural network which was retrained for human emotion classification. A variety of datasets, as well as our own image dataset, is used to train the model. Humans are capable of producing thousands of facial actions during communication that vary in complexity, intensity, and meaning. This paper analyses the limitations with existing system emotion recognition. This project achieved 85 percent accurate results. Finally, a live video stream connected to a face detector feeds images to the neural network. The network subsequently classifies an arbitrary number of faces per image simultaneously in real time. Purposed system depends upon human face as we know face reflects the emotions. The results demonstrate the feasibility of implementing neural networks in real time to detect human emotion.

Keywords: convolutional neural network, feature extraction, Emotion

TABLE OF CONTENT

CHAPTER NO.	TITLE	PAGE NO.
	ABSTRACT	i
	LIST OF ABBREVIATIONS	ii
	LIST OF FIGURES	iii
1	INTRODUCTION	9
	1.1 INTRO	9
	1.2 HUMAN EMOTION	9
	1.3 IMAGE CLASSIFICATION TECHNIQUES	10
	1.4 DATASET	10
	1.5 TOOLS AND PLATFORM	11
2	BACKGROUND	12
	2.1 INTRO	12
	2.2 EVOLUTION TIMELINE	12
	2.3 NEED OF FEATURE EXTRACTION	12
	2.4 NEED FOR EMOTION CLASSIFICATION	13
3	IMAGE FEATURE	14
	3.1 FACTS	14
	3.2 LANDMARKS	15

	3.3 FEATURE DESCRIPTORS	15
4	TOOLS AND LIBRARIES USED	17
	4.1 OPENCV	17
	4.2 DLIB	17
	4.3 PYTHON	17
	4.4 SCIKIT LEARN	18
	4.5 JUPYTER NOTEBOOK	18
	4.6 DATABASE	18
5	RESULT AND CONCLUSIONS	20
	5.1 RESULT	20
	5.2 CONCLUSION	21
6	REFERENCES	22
7	APPENDIX	23
	A. SOURCE CODE	24

LIST OF ABBREVIATIONS:-

ABBREVIATION

AI
ANN
CV

MEANING

Artificial Intelligence
Artificial Neural Network
Computer Vision

CHAPTER 1

INTRODUCTION

1.1 INTRO

Ever since computers were developed, scientists and engineers thought of artificially intelligent systems that are mentally and/or physically equivalent to humans. In the past decades, the increase of generally available computational power provided a helping hand for developing fast learning machines, whereas the internet supplied an enormous amount of data for training. These two developments boosted the research on smart self-learning systems, with neural networks among the most promising techniques.

1.2 HUMAN EMOTION

A key feature in human interaction is the universality of facial expressions and body language. Already in the nineteenth century, Charles Darwin published upon globally shared facial expressions that play an important role in non-verbal communication. In 1971, Ekman and Friesen declared that facial behaviors are universally associated with particular emotions. Apparently humans, but also animals, develop similar muscular movements belonging to a certain mental state, despite their place of birth, race, education, etcetera. Hence, if properly modelled, this universality can be a very convenient feature in human-machine interaction: a well trained system can understand emotions, independent of who the subject is. One should keep in mind that facial expressions are not necessarily directly translatable into emotions, nor vice versa. Facial expression is additionally a function of e.g. mental state, while emotions are also expressed via body language and voice. More elaborate emotion recognition systems should therefore also include these latter two contributions. However, this is out of the scope of this research and will remain a recommendation for future work. Readers interested in research on emotion classification via speech recognition are referred to Nicholson et al. As a final point of attention, emotions should not be confused with mood, since mood is considered to be a long-term mental state. Accordingly, mood recognition often involves longstanding analysis of someone's behaviour and expressions, and will therefore be omitted in this work.

1.3 IMAGE CLASSIFICATION TECHNIQUES

The growth of available computational power on consumer computers in the beginning of the twenty-first century gave a boost to the development of algorithms used for interpreting pictures. In the field of image classification, two starting points can be distinguished. On the one hand pre-programmed feature extractors can be used to analytically break down several elements in the picture in order to categorize the object shown. Directly opposed to this approach, self-learning neural networks provide a form of blackbox identification technique. In the latter concept, the system itself develops rules for object classification by training upon labelled sample data. An extensive overview of analytical feature extractors and neural network approaches for facial expression recognition is given by Fasel and Luetten [6]. It can be concluded that by the time of writing, at the beginning of the twenty-first century, both approaches work approximately equally well. However, given the current availability of training data and computational power it is the expectation that the performance of neural network based models can be significantly improved by now.

1.4 DATASET

Neural networks, and deep networks in particular, are known for their need for large amounts of training data. Moreover, the choice of images used for training are 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 non-frontal faces and pictures with the label contemptuous are taken out of the RaFD data, since these are not represented in the FEREC-2013 training set. Furthermore, with use of the Haar Feature-Based Cascaded Classifier inside the OpenCV framework all data is preprocessed. For every image, only the square part containing the face is taken, rescaled, and converted to an array with 48x48 grey-scale values.

1.5 TOOLS & PLATFORM

Minimum Requirement:-

Processor:- Pentium 4/1.2 Ghz Celeron/Duron process or

CPU Speed:- 1.4 GHz

RAM:- 512 MB

Hard Disk:- 80 GB

OS:- Windows 98/2000/XP

Mouse:-

Keyboard:-

Sound Card:- NO

Language:- Java 5.0

Server:- Apache Tomcat 4.1

CHAPTER 2

BACKGROUND

2.1 INTRO

This section focuses on the background and importance of facial expressions for sentiment analysis. It is bifurcated into four subsections. Firstly, we discuss the evolution of the time- line of facial recognition methods. Secondly, we discuss the need for Facial Detection, Dimension Reduction, and Normalization for sentiment analysis. In the third subsection, we focus on the need for feature extraction from the face image. Finally, we highlight the need for emotion classification

2.2 EVOLUTION TIMELINE

Figure 3 gives a brief overview on the evolutionary time- line of facial sentimental recognition methods given by the researchers across the globe along with the datasets. There exists various algorithms for FER such as traditional state- of-the-art algorithms and DL-based algorithms proposed by various researchers till 2020. The emotion recognition was first stated in the paper proposed by Bassili *et al.* in 1978 where authors have classified the emotions into six basic gestures such as happiness, sadness, fear, surprise, anger, and disgust. Different algorithms (traditional and DL) were used for FER by the authors. For example, Padgett *et al.*

2.3 NEED FOR FEATURE EXTRACTION

Facial Expression analysis comprises of various methods such as facial landmark identification, feature extraction, and different feature extraction databases. Facial landmarks are drawn by the facial key points which are derived from the geometry of the face [16]. Feature Extraction is done after preprocessing phase [49]. There are two methods available for feature extraction are appearance-based extraction and geometric-based extraction. The geometric-based method extracts feature like edge features and corner features. Verma *et al.* [50] analyzed the performance of the feature extraction technique Gabor filter. They also tested the average gabor filter and compared both the filtering techniques to enhance the recognition rate.

The second method, which is an appearance-based method, takes care of the states of

different points of the face, such as the position of the eye, shape of important points such as mouth and eyebrows using the salient point features. The majority of the traditional methods have used Local Binary Pattern (LBP) as the feature extraction technique, which is a generic-based framework for the extraction of features from the static image. It converts the most important features of the input image, as mentioned above, into a histogram.

2.4 NEED FOR EMOTION CLASSIFICATION

The third step in the FER is the Emotion Classification. There are various methods that are used for the classification of emotions after applying face detection and feature extraction algorithms. The various classification algorithms are convolutional neural network (CNN) [53], SVM, and restricted boltzmann machine (RBM). The most widely used method for classification is CNN. It is the most efficient algorithm as it can be applied directly to the input image without applying any feature extraction and face detection algorithms and still gets better accuracy over the input data [54]. The number of images in the training data set also has a huge impact on classification results. CNN faces a huge challenge in the training of limited-image dataset. So, the models which are built on a limited dataset can use the SVM algorithm for feature extraction and face detection. The emotions of a human are not static, it varies time-to-time. So, the classification of situation-based emotions is challenging.

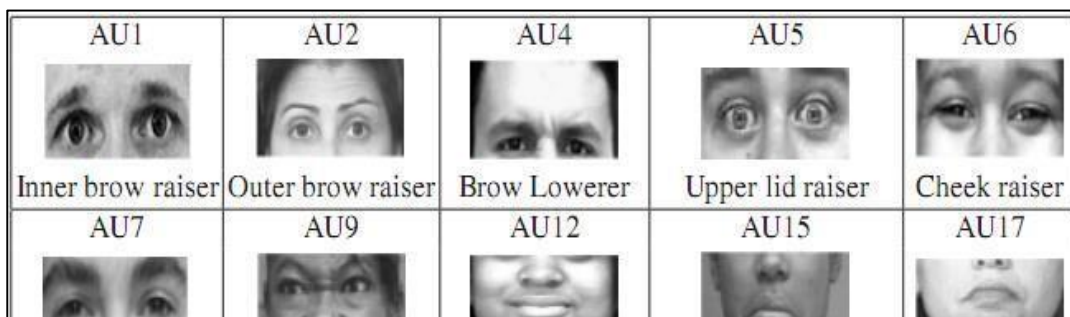
CHAPTER 3

IMAGE FEATURES

We can derive different types of features from the image and normalize it in vector form. We can employ various types of techniques to identify the emotion like calculating the ellipses formed on the face or the angles between different parts like eyes, mouth etc. Following are some of the prominent features which can be used for training machine learning algorithms:

3.1 FACTS

Facial Action Coding System is used to give a number to facial moment. Each such number is called as action unit. Combination of action units result in a facial expression. The micro changes in the muscles of the face can be defined by an action unit. For example, a smiling face can be defined in terms of action units as 6 + 12, which simply means movement of AU6 muscle and AU12 muscle results in a happy face. Here Action Unit 6 is cheek raiser and Action Unit 12 is lip corner puller. Facial action coding system based on action units is a good system to determine which facial muscles are involved in which expression. Real time face models can be generated based on them.



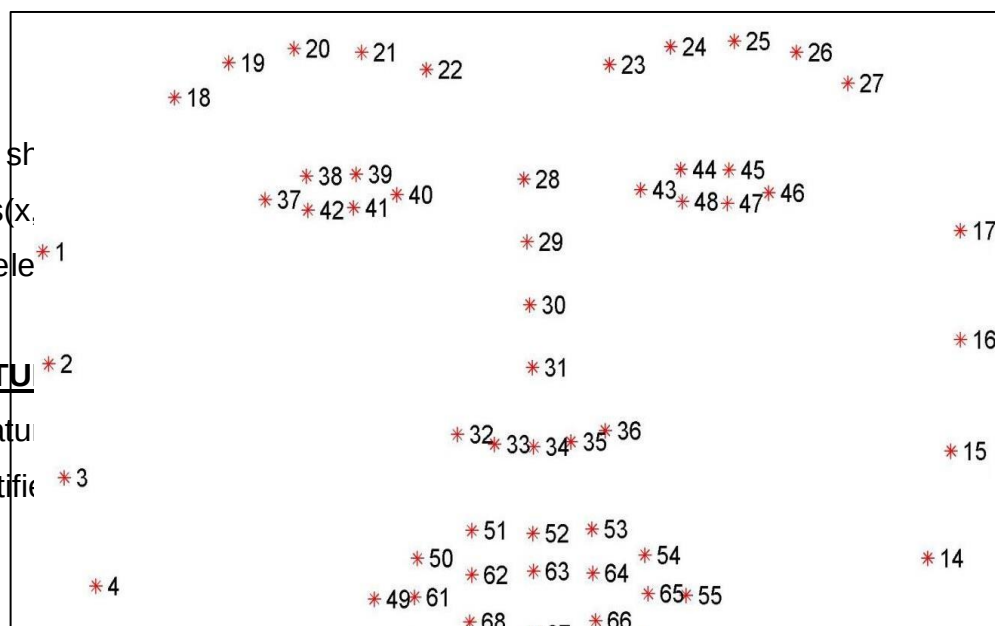
3.2 LANDMARKS

Landmarks on the face are very crucial and can be used for face detection and recognition. The same landmarks can also be used in the case of expressions. The Dlib library has a 68 facial landmark detector which gives the position of 68 landmarks on the face.

Figure 2 shows the coordinates (x, y) of the 68 facial landmarks on a face. The landmarks are numbered 1 to 68, with 1 at the top left and 68 at the bottom right. The landmarks are distributed across the face, with specific areas like the eyes, nose, mouth, and chin being more densely marked.

3.3 FEATURES

Good features are identified from the face images. The features are used to detect the edges in the face images.



images, we have many feature detector algorithms in the OpenCV library such as Harris corner detector.

These feature detectors take into account many more factors such as contours, hull and convex. The Key-points are corner points or edges detected by the feature detector algorithm. The feature descriptor describes the area surrounding the key-point. The description can be anything including raw pixel intensities or co-ordinates of the surrounding area. The key-point and descriptor together form a local feature. One example of a feature descriptor is a histogram of oriented gradients. ORB (based on BRIEF), SURF, SIFT etc. are some of the feature descriptor algorithms.

CHAPTER 4

TOOLS AND LIBRARIES USED

The logic is the same as backtracking, we go cell by cell instead of row by row. That is, “try every possible number for the next empty cell. If you find a solution with the current number, stop. If you don't, then try the next number.”

4.1 OPEN CV

OpenCV is the library we will be using for image transformation functions such as converting the image to grayscale. It is an open source library and can be used for many image functions and has a wide variety of algorithm implementations. C++ and Python are the languages supported by OpenCV. It is a complete package which can be used with other libraries to form a pipeline for any image extraction or detection framework. The range of functions it supports is enormous, and it also includes algorithms to extract feature descriptors.

4.2 DLIB

Dlib is another powerful image-processing library which can be used in conjunction with Python, C++ and other tools. The main function this library provides is of detecting faces, extracting features, matching features etc. It has also support for other domains like machine learning, threading, GUI and networking.

4.3 PYTHON

Python is a powerful scripting language and is very useful for solving statistical problems involving machine learning algorithms. It has various utility functions which help in pre-processing. Processing is fast and it is supported on almost all platforms. Integration with C++ and other image libraries is very easy, and it has in-built functions and libraries to store and manipulate data of all types. It provides the pandas and numpy framework which helps in manipulation of data as per our need. A good feature set can be created using the numpy arrays which can have n-dimensional data.

4.4 SCIKIT- LEARN

Scikit-learn is the machine learning library in python. It comprises of matplotlib, numpy and a wide array of machine learning algorithms. The API is very easy to use and understand. It has many functions to analyze and plot the data. A good feature set can be formed using many of its feature reduction, feature importance and feature selection functions. The algorithm it provides can be used for classification and regression problems and their sub-types.

4.5 Jupyter Notebook

Jupyter Notebook is the IDE to combine python with all the libraries we will be using in our implementation. It is interactive, although some complex computations require time to complete. Plots and images are displayed instantly. It can be used as a one stop for all our requirements, and most of the libraries like Dlib, OpenCV, Scikit-learn can be integrated easily.

4.6 Database

We have used the extended Cohn-Kanade database (CK+) and Radboud Faces database (RaFD). CK+ has around 593 images for 123 subjects. Only 327 files have labeled/identified emotions. It covers all the basic human emotions displayed by the face. The emotions and codes are as follows: 1 – Angry, 2 – Contempt, 4 – Fear, 5 – Happy, 6 – Sadness, 7 – Surprise. The database is widely used for emotion detection research and analysis. There are 3 more folders along with the images. FACS contains action units for each image. Landmark contains AAM tracked facial features of the 68 facial points. Emotion contains emotion label for the 327 files. Radboud Faces Database [19] is a standard database having equal number of files for all emotions. It has images of 67 subjects displaying 8 emotions: Neutral included. The pictures are taken in 5 different

camera poses. Also, the gaze is in 3 directions. We are using only front facing images for our experiment. We have a total of 536 files with 67 models displaying 8 different emotions.

CHAPTER 5

RESULT AND CONCLUSIONS

5.1 Result

We applied support vector machines to our dataset and predicted the results. The results were interpreted using confusion matrix and accuracy metric. The train:test split was 75:25. We also did cross-validation on the dataset to remove any biases. Value of split was chosen as 4 because the resultant splits will have same number of images as our 25% test set. In our experiment SVM with linear kernel performed better than other kernels. Rbf gave us the worst performance, whereas poly was as good as linear kernel. We tried to keep the test set % same for both split and cross-validated data so as to have uniformity in results. The mean cross-validation score was also approximately equal to the accuracy score achieved by the split. Figure 17 shows the heat-map of the confusion matrix from our multi-class classification results. On further analysis of the confusion matrix, we see that the diagonals have higher weights; there are a few misclassifications in every class except class 4: Fear. Many of the samples were misclassified into other classes such as contempt, disgust, fear and sadness. Fear was detected accurately for all the samples. The samples for disgust and sadness were misclassified as anger in some cases. This can be due to the similarity of features between these emotions. For our next part, we considered all the 68 facial point features and determined which of these features actually help to determine the emotion. Jaw was not adding any difference in all of the micro-expressions. Hence, we decided to ignore the jaw facial landmark positions and continue with eyes, eyebrows, mouth and nose for our analysis. Our feature set was still the same with dimensions $327 \times 68 \times 2$, but all the jaw facial landmarks had co-ordinates as 0. Also, we kept the split ratio of test-train dataset same as above. Following are the results for this experiment. Our focus is on the cross-validation score as it takes entire database into consideration for training, thus removing any biases. After removing jaw-line features the cross-validation results also increased from 78% to 80%.

To improve the results, we did some fine tuning. We changed the data split to 70:30 and cross validation folds to 5. We use stratified sampling for the split. We also extracted more features from the available 68 facial point landmarks.

5.2 Conclusion

Our implementation can roughly be divided into 3 parts:

- 1) Face detection
- 2) Feature extraction
- 3) Classification using machine learning algorithms

Feature extraction was very important part of the experiment. The added distance and area features provided good accuracy for CK+ database (89%). But for cross-database experiment we observed that raw features worked best with Logistic Regression for testing RaFD database and Mobile images dataset. The accuracy was 66% and 36% for both using CK+ dataset as training set. The additional features (distance and area) reduced the accuracy of the experiment for SVM.

The algorithm generalized the results from the training set to the testing set better than SVM and other algorithms. The results of the emotion detection algorithm gave average accuracy up to 86% for RaFD database and 87% for CK+ database for cross-validation=5. RaFDdataset had equal number of classes; hence, cross-validation did not help in improving the

CHAPTER 6

REFERENCES

- [1] Neerja and E. Walia, "Face Recognition Using Improved Fast PCA Algorithm," *2008 Congress on Image and Signal Processing*, Sanya, Hainan, 2008
- [2] H. Ebine, Y. Shiga, M. Ikeda and O. Nakamura, "The recognition of facial expressions with automatic detection of the reference face," *2000 Canadian Conference on Electrical and Computer Engineering. Conference Proceedings. Navigating to a New Era (Cat. No.00TH8492)*, Halifax, NS, 2000, pp. 1091-1099 vol.2.
- [3] A. C. Le Ngo, Y. H. Oh, R. C. W. Phan and J. See, "Eulerian emotion magnification for subtle expression recognition," *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Shanghai, 2016
- [4] V. Kazemi and J. Sullivan, "One millisecond face alignment with an ensemble of regression trees," *2014 IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, 2014
- [5] M. Dahmane and J. Meunier, "Emotion recognition using dynamic grid-based HoG features," *Face and Gesture 2011*, Santa Barbara, CA, 2011
- [6] K. M. Rajesh and M. Naveenkumar, "A robust method for face recognition and face emotion detection system using support vector machines," *2016 International Conference on Electrical, Electronics, Communication, Computer and Optimization Techniques (ICEECCOT)*, Mysuru, 2016
- [7] C. Loconsole, C. R. Miranda, G. Augusto, A. Frisoli and V. Orvalho, "Real-time emotion recognition novel method for geometrical facial features extraction," *2014 International Conference on Computer Vision Theory and Applications (VISAPP)*, Lisbon, Portugal, 2014
- [8] J. M. Saragih, S. Lucey and J. F. Cohn, "Real-time avatar animation from a single image," *Face and Gesture 2011*, Santa Barbara, CA, USA, 2011
- [9] G. T. Kaya, "A Hybrid Model for Classification of Remote Sensing Images

With Linear SVM and Support Vector Selection and Adaptation," in *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 6, no. 4, pp. 1988-1997, Aug. 2013

[10] X. Jiang, "A facial expression recognition model based on HMM," Proceedings of 2011 International Conference on Electronic & Mechanical Engineering and Information Technology, Harbin, Heilongjiang, China, 2011

[11] J. J. Lee, M. Zia Uddin and T. S. Kim, "spatiotemporal human facial expression recognition using fisher independent component analysis and Hidden Markov Model," *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Vancouver, BC, 2008

[12] Xiaoxu Zhou, Xiangsheng Huang, Bin Xu and Yangsheng Wang, "Real-time facial expression recognition based on boosted embedded hidden Markov model," *Image and Graphics (ICIG'04), Third International Conference on*, Hong Kong, China, 2004

CHAPTER 7

A. SOURCE CODE

```
from fer import Video
from fer import FER
import os
import sys
import pandas as pd

# Put in the location of the video file that has to be processed
location_videofile = "/content/Video_One.mp4"

# Build the Face detection detector
face_detector = FER(mtcnn=True)
# Input the video for processing
input_video = Video(location_videofile)

# The Analyze() function will run analysis on every frame of the input video.
# It will create a rectangular box around every image and show the emotion values next to that.
# Finally, the method will publish a new video that will have a box around the face of the human
# with live emotion values.
processing_data = input_video.analyze(face_detector, display=False)

# We will now convert the analysed information into a dataframe.
# This will help us import the data as a .CSV file to perform analysis over it later
vid_df = input_video.to_pandas(processing_data)
vid_df = input_video.get_first_face(vid_df)
vid_df = input_video.get_emotions(vid_df)

# Plotting the emotions against time in the video
pltfig = vid_df.plot(figsize=(20, 8), fontsize=16).get_figure()

# We will now work on the dataframe to extract which emotion was prominent in the video
angry = sum(vid_df.angry)
disgust = sum(vid_df.disgust)
fear = sum(vid_df.fear)
happy = sum(vid_df.happy)
sad = sum(vid_df.sad)
surprise = sum(vid_df.surprise)
neutral = sum(vid_df.neutral)

emotions = ['Angry', 'Disgust', 'Fear', 'Happy', 'Sad', 'Surprise', 'Neutral']
emotions_values = [angry, disgust, fear, happy, sad, surprise, neutral]
```

```
score_comparisons = pd.DataFrame(emotions, columns = ['Human Emotions'])  
score_comparisons['Emotion Value from the Video'] = emotions_values  
score_comparisons
```