AI DESKTOP PARTNER

A Report for the Evaluation 3 of Project 2

Submitted By

RAM LAKHAN

1613101554

In partial Fulfillment for the award of the degree

Of

Bachelor Of Technology

IN

Computer Science and Engineering

SCHOOL OF COMPUTER SCIENCE AND ENGINEERING

Under the Supervision Of

DR. N Thillaiarasu

APRIL/MAY-2020

TABLE OF CONTENTS

ABSTRACT

A desktop partner bot who chats with you when you are bored. The bot is built with an Machine Learning algorithm. It chats with you as a real person with amusing replies which doesn't make the user know he is really talking to a bot. The bot is built with a limited dictionary but uses a great algorithm to imitate a real person. The bot can be used to find you an amusing partner and help you in bad times. The desktop partner comes with a real life person imitation (RLPI) System designed in 2013 for putting forward a real time intelligent chatting session for users. The algorithm used here has a intelligently built in logic and is designed to better chat with Indians since it's dictionary is better configured as per Indian mentality.

Emotion recognition plays very important role in recent days to improve both openness and effectiveness of Human-computer interaction. Emotions include the interpretation, perception and response of the feelings related to the experience of any particular situation. Emotion recognition is the task of recognizing a person's emotional state such as anger, sad, happy, etc. Emotion recognition consist of the classification of emotion from different approaches such as face and body pose of person. The applications of emotion recognition are monitoring, law, entertainment, e- learning, medicine and many others. In this paper, we present a new approach for building desktop application for chat bot using gestures.For example, if the user is feeling sad, then system will automatically fetch a joke from database and send it to the user on the window terminal. The system is able to make a conversation through chatting application. System is able to send some links, web pages or information by recognizing response from user. For this whole system, we are using technologies like ML,Image Processing, Data mining , Python, CNN etc.

INTRODUCTION

1.1 Overall Description:

Mood of person describes the inherent emotional meaning. Detection of mood play a significant role in human dialogue. The Chat bot is built using an artificial intelligence algorithm. The bot chats with you as a real person,with entertaining replies which doesn't make the user know he is really talking to a bot. In proposed system, we are presenting a new approach for building application for chat bot using gestures . My application recognizes the mood of a user through camera. For example, if the user is feeling sad, then system will automatically fetch a sad songs and from database and send it to the user on the window terminal. The system is able to make a conversation through chatting application. System is able to send some links, web pages or information by recognizing response from user. The human body is premeditated to know-how stress and reacts to it. It chats with you as a real person with amusing replies which doesn't make the user know he is really talking to a bot. The bot is built with a limited dictionary but uses a great algorithm to imitate a real person. The bot can be used to find you an amusing partner and help you in bad times. The AI desktop partner comes with a real life person imitation (RLPI) System designed in 2013 for putting forward a real time intelligent chatting session for users.

1.2 Backgrounds:

One of the current top applications of artificial intelligence using neural networks is the **recognition of faces** in photos and videos. Most techniques process visual data and search for general patterns present in human faces. Face recognition can be used for surveillance purposes by law enforcers as well as in crowd management. Other present-day applications involve automatic blurring of faces on Google Street view footage and automatic recognition of Facebook friends in photos.

An even more advanced development in this field is **emotion recognition**. In addition to only identifying faces, the computer uses the arrangement and shape of e.g. eyebrows and lips to

determine the facial expression and hence the emotion of a person. One possible application for this lies in the area of surveillance and behavioural analysis by law enforcement. Furthermore such techniques are used in digital cameras to automatically take pictures when the user smiles. However, the most promising applications involve the humanization of artificial intelligent systems. If computers are able to keep track of the mental state of the user, robots can react upon this and behave appropriately. Emotion recognition therefore plays a key-role in improving human machine interaction.

1.3 Objective:

In this research we mainly focus on neural network based artificially intelligent systems capable of deriving the emotion of a person through pictures of his or her face. Different approaches from existing literature will be experimented with and the results of various choices in the design process will be evaluated. The main research question therefore reads as follows: **How can an artificial neural network be used for interpreting the facial expression of a human?**

The remainder of this article describes the several steps taken to answer the main research question, i.e. the sub-questions. In section 2, a literature survey will clarify *what the role of facial expressions is in emotion recognition* and *what types of networks are suitable for automated image classification*. The third section explains *how the neural networks under consideration are structured* and *how the networks are trained*. Section 4 describes *how the final model performs* after which a conclusion and some recommendations follow in the last section. It may be noted that the aim of our work is not to design an emotion recognizer from scratch but rather to review design choices and enhance existing techniques with some new ideas.

EXISTING SYSTEM

For the development of a system that is able to recognize emotions through facial expressions, previous research on the way humans reveal emotions as well as the theory of automatic image categorization is reviewed. In the first part of this section, the role of interpreting facial expressions in emotion recognition will be discussed. The latter part surveys previous studies on automatic image classification.

2.1 Human emotions:

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 [3]. In 1971, Ekman & Friesen declared that *facial behaviors are universally associated with particular emotions* [5]. 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 humanmachine 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 [6]. 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. [14]. 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.

2.2 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 ’black box’ 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 Luettin [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. Some recent achievements will be listed below.

(i) A breakthrough publication on automatic image classification in general is given by Krizhevsky and Hinton [9]. This work shows a deep neural network that resembles the functionality of the human visual cortex. Using a self-developed labelled collection of 60000 images over 10 classes, called the CIFAR-10 data set, a model to categorize objects from pictures is obtained. Another important outcome of the research is the visualization of the filters in the network, such that it can be assessed how the model breaks down the pictures.

(ii) In another work which adopts the CIFAR-10 data set [2], a very wide and deep network architecture is developed, combined with GPU support to decrease training time. On popular data sets, such as the MNIST handwritten digits, Chinese characters, and the CIFAR-10 images, near-human performance is achieved. The extremely low error rates beat prior state-of-theart

results significantly. However it has to be mentioned that the network used for the CIFAR10 data set consists of 4 convolutional layers with 300 maps each, 3 max pooling layers, and 3 fully connected output layers. As a result, although a GPU was used, the training time was several days.

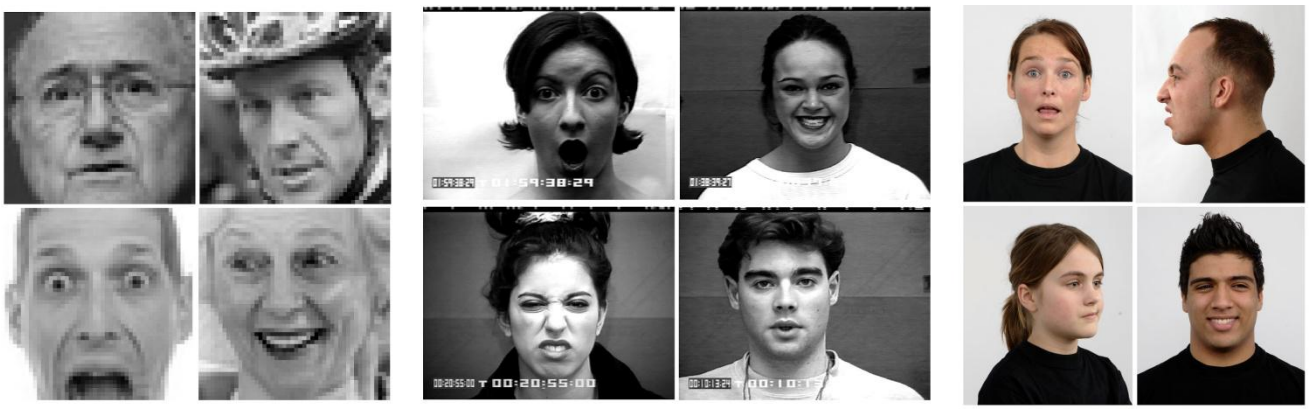
(iii) In 2010, the introduction of the yearly Imagenet challenge [4] boosted the research on image classification and the belonging gigantic set of labelled data is often used in publications ever since. In a later work of Krizhevsky et al. [10], a network with 5 convolutional, 3 max pooling, and 3 fully connected layers is trained with 1.2 million high resolution images from the ImageNet LSVRC-2010 contest. After implementing techniques to reduce overfitting, the results are promising compared to previous state-of-theare models. Furthermore, experiments are done with lowering the network size, stating that the number of layers can be significantly reduced while the performance drops only a little.

(iv) With respect to facial expression recognition in particular, Lv et al. [13] present a deep belief network specifically for use with the Japanese Female Facial Expression (JAFFE) and extended Cohn-Kanade (CK+) databases. The most notable feature of the network is the hierarchical face parsing concept, i.e. the image is passed through the network several times to first detect the face, thereafter the eyes, nose, and mouth, and finally the belonging emotion. The results are comparable with the accuracy obtained by other methods on the same database, such as Support Vector Machine (SVM) and Learning Vector Quantization (LVQ).

(v) Another work on the Cohn-Kanade database [1] makes use of Gabor filtering for image processing and Support Vector Machine (SVM) for classification. A Gabor filter is particularly suitable for pattern recognition in images and is claimed to mimic the function of the human visual system. The emotion recognition accuracy are high, varying from 88% on anger to 100% on surprised. A big disadvantage of the approach however is that very precise pre-processing of

the data is required, such that every image complies to a strict format before feeding it into the classifier.

(vi) One of the most recent studies on emotion recognition describes a neural network able to recognize race, age, gender, and emotion from pictures of faces [7]. The dataset used for the latter category is originating from the Facial Expression Recognition Challenge (FERC-2013). A clearly organized deep network consisting of 3 convolutional layers, 1 fully connected layer, and some small layers in between obtained an average accuracy of 67% on emotion classification, which is equal to previous state-of-the-art publications on the same dataset. Furthermore this thesis lays down a valuable analysis of the effect of adjusting the network size, pooling, and dropout. Underlined by some other literature, the most promising concept for facial expression analysis is the use of deep convolutional neural networks. However, the network from [2] (ii) is considered to be too heavy for our limited amount of available processing resources. The original network from [10] (iii) is large as well, but smaller versions are claimed to be equally suitable. Furthermore, due to their somewhat analytical and unconventional approaches, we will not evaluate [13] (iv) and [1] (v). Hence, in the next section, three deep architectures in total will be subjected to an emotion classification problem. These architectures are derived from, but not necessarily equal to, the networks described at items i, iii, and vi.



PROPOSED SYSTEM

Proposed system uses a algorithms and technologies such as Haar cascade detection. In the system pictures were taken and according to that mood get detected. Inputs like face and emotions are taken from picture, and system will be providing chat box to give response.

Proposed system uses Machine Learning techniques to provide Chabot which provides chatting interface to communicate with user. System present a new approach for building desktop application for chat bot using text and gestures. The system is able to make a conversation through chatting application. System is able to send some links, web pages or information by recognizing response from user.Our Proposed system Detects smile and stress. If smile is detected than jokes pop-ups will be shown on the screen, if stress is detected than inspirational quotes pop-ups will be shown on the screen. On detection of smile happy songs are played. On detection of stress inspirational songs are played.

The algorithm used here has a perceptively built in logic and is designed to better chat with Indians since its dictionary is better constructed as per Indian approach. A desktop partner bot who chats with you when you are tired. The bot is built with an artificial intelligence process. It chats with you as an actual person with amusing answers which doesn’t make the user know he is really talking to a bot.

This proposed system based on facial expression extracted will generate a playlist automatically thereby reducing the effort and time involved in rendering the process manually. Thus the proposed system tends to reduce the computational time involved in obtaining the results and the overall cost of the designed system, thereby increasing the overall accuracy of the system. Facial expressions are captured using an inbuilt camera.

TECHNOLOGIES TO BE USED

1. Introduction to Image Processing:

In order to get an enhanced image and to extract some useful information out of it, the method of Image Processing can be used. It is a very efficient way through which an image can be converted into its digital form subsequently performing various operations on it. This is a technique similar to signal processing, in which the input given is a 2D image, which is a collection of numbers ranging from 0 to 255 which denotes the corresponding pixel value.

2. Introduction to OpenCV:

OpenCV is one of the most popular computer vision libraries. If you want to start your journey in the field of computer vision, then a thorough understanding of the concepts of OpenCV is of paramount importance. In this article, I will try to introduce the most basic and important concepts of OpenCV in an intuitive manner.

OpenCV is a cross-platform library using which we can develop real-time computer vision applications. It mainly focuses on image processing, video capture and analysis including features like face detection and object detection.

3. Machine Learning:

Machine learning has solved many problems by choosing one machine learning algorithm, feeding in data, and getting the result. We do not need to build our own neural network. We have access to a trained model that can be used. It does exactly what we need it to do — outputs a bunch of numbers (face encodings) when we pass in the image of someone's face; comparing face encodings of faces from different images will tell us if someone's face matches with anyone we have images of.

4. **OpenCV-Python:**

Python is a general purpose programming language started by **Guido van Rossum**, which became very popular in short time mainly because of its simplicity and code readability. It enables the programmer to express his ideas in fewer lines of code without reducing any readability.

Compared to other languages like C/C++, Python is slower. But another important feature of Python is that it can be easily extended with C/C++. This feature helps us to write computationally intensive codes in C/C++ and create a Python wrapper for it so that we can use these wrappers as Python modules. This gives us two advantages: first, our code is as fast as original C/C++ code (since it is the actual C++ code working in background) and second, it is very easy to code in Python. This is how OpenCV-Python works, it is a Python wrapper around original C++ implementation.

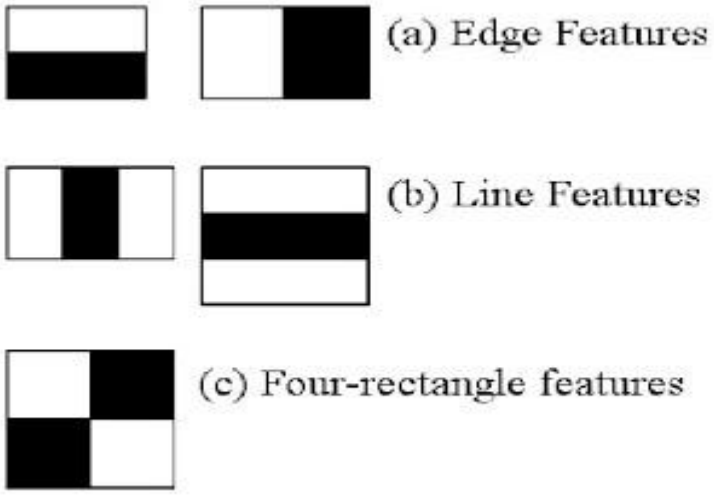
And the support of Numpy makes the task more easier. **Numpy** is a highly optimized library for numerical operations. It gives a MATLAB-style syntax. All the OpenCV array structures are converted to-and-from Numpy arrays. So whatever operations you can do in Numpy, you can combine it with OpenCV, which increases number of weapons in your arsenal. Besides that, several other libraries like SciPy, Matplotlib which supports Numpy can be used with this.

So OpenCV-Python is an appropriate tool for fast prototyping of computer vision problems.

5. **Face Detection using Haar Cascades:**

Object Detection using Haar feature-based cascade classifiers is an effective object detection method proposed by Paul Viola and Michael Jones in their paper, “Rapid Object Detection using a Boosted Cascade of Simple Features” in 2001. It is a machine learning based approach where a cascade function is trained from a lot of positive and negative images. It is then used to detect objects in other images.

Here we will work with face detection. Initially, the algorithm needs a lot of positive images (images of faces) and negative images (images without faces) to train the classifier. Then we need to extract features from it. For this, haar features shown in below image are used. They are just like our convolutional kernel. Each feature is a single value obtained by subtracting sum of pixels under white rectangle from sum of pixels under black rectangle.



Now all possible sizes and locations of each kernel is used to calculate plenty of features. (Just imagine how much computation it needs? Even a 24x24 window results over 160000 features). For each feature calculation, we need to find sum of pixels under white and black rectangles. To solve this, they introduced the integral images. It simplifies calculation of sum of pixels, how large may be the number of pixels, to an operation involving just four pixels. Nice, isn't it? It makes things super-fast.

6**. CNN :**

A Convolutional Neural Network (ConvNet/CNN) is a Deep Learning algorithm which can take in an input image, assign importance (learnable weights and biases) to various aspects/objects in the image and be able to differentiate one from the other. The pre-processing required in a ConvNet is much lower as compared to other classification algorithms. While in primitive methods filters are hand-engineered, with enough training, ConvNets have the ability to learn these filters/characteristics.

The architecture of a ConvNet is analogous to that of the connectivity pattern of Neurons in the Human Brain and was inspired by the organization of the Visual Cortex. Individual neurons respond to stimuli only in a restricted region of the visual field known as the Receptive Field. A collection of such fields overlap to cover the entire visual area.

7. **Introduction to the Python Deep Learning Library TensorFlow:**

TensorFlow is an end-to-end open source platform for machine learning. It has a comprehensive, flexible ecosystem of tools, libraries and community resources that lets researchers push the state-of-the-art in ML and developers easily build and deploy ML powered applications.

TensorFlow is a Python library for fast numerical computing created and released by Google.

It is a foundation library that can be used to create Deep Learning models directly or by using wrapper libraries that simplify the process built on top of Tensor flow.

What is Tensor Flow?

Tensor Flow is an open source library for fast numerical computing.

It was created and is maintained by Google and released under the Apache 2.0 open source license. The API is nominally for the Python programming language, although there is access to the underlying C++ API.

Unlike other numerical libraries intended for use in Deep Learning like Theano, Tensor Flow was designed for use both in research and development and in production systems, not least Rank Brain in Google and the fun Deep Dream project.

It can run on single CPU systems, GPUs as well as mobile devices and large scale distributed systems of hundreds of machines.

ARCHITECTURE DIAGRAMS

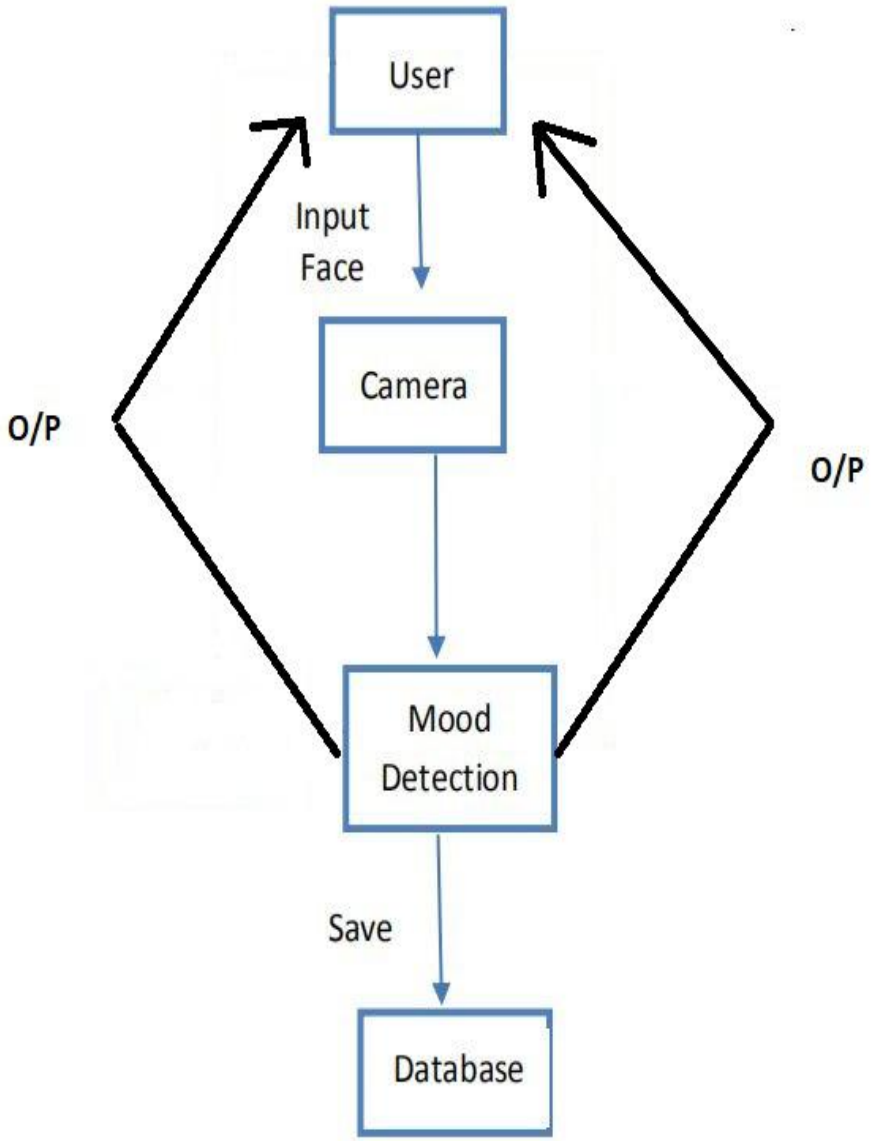


Fig1. Block diagram of proposed system

OUTPUT

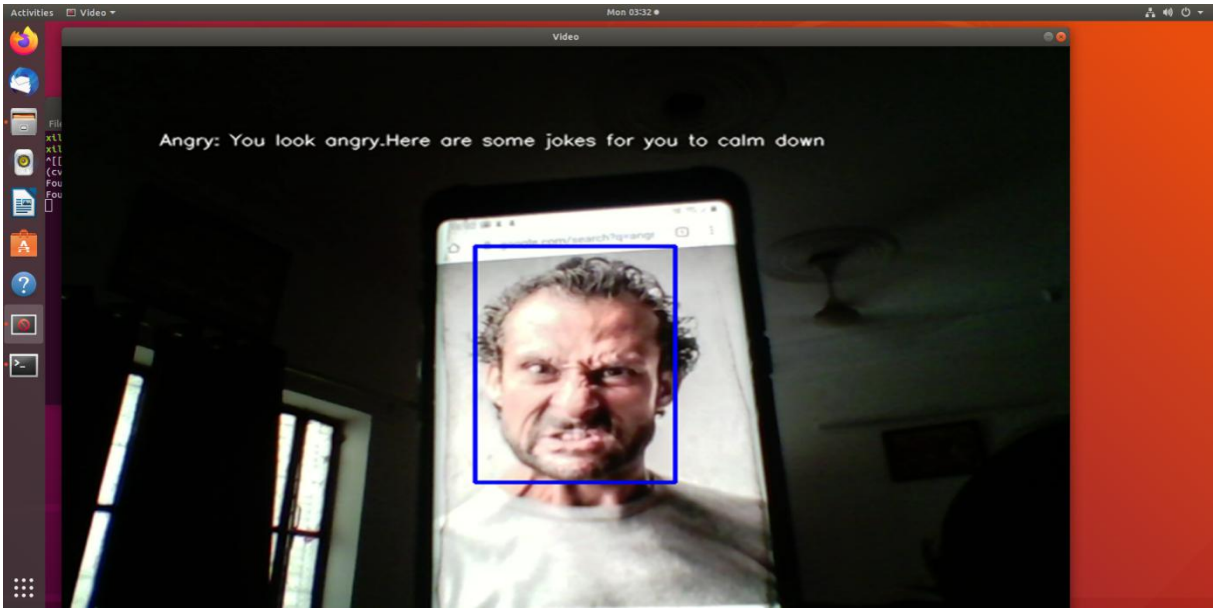


Fig 1.ANGRY

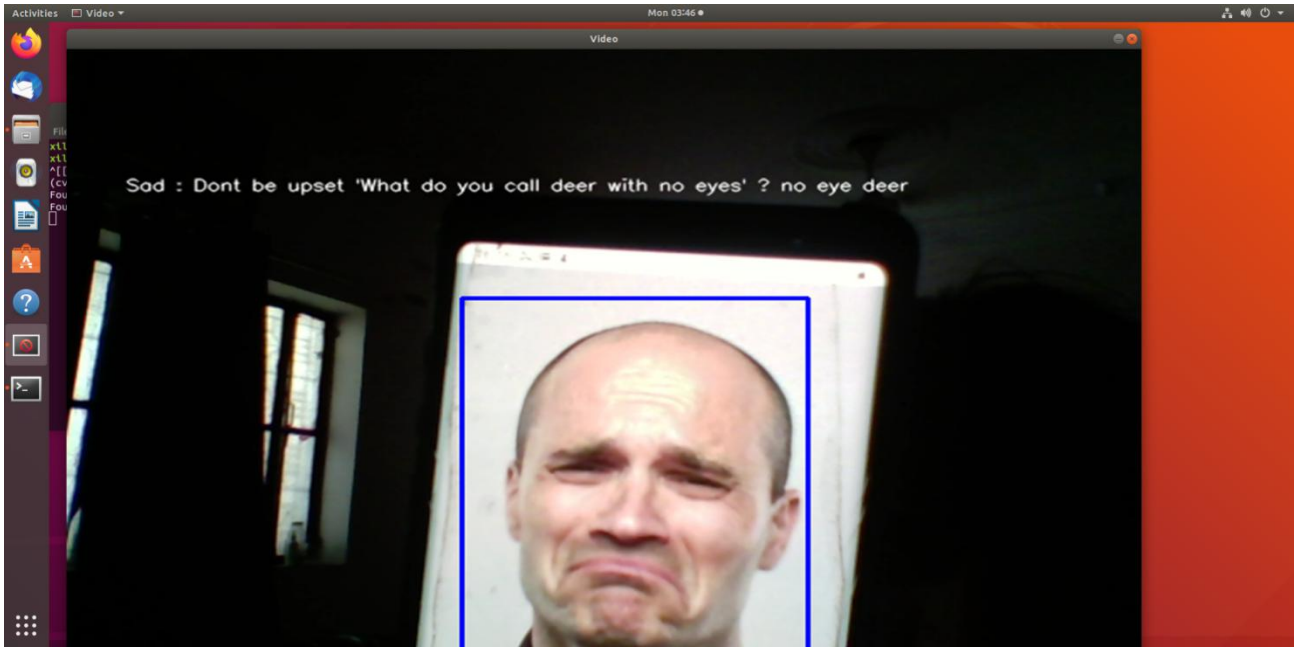


Fig 2.SAD

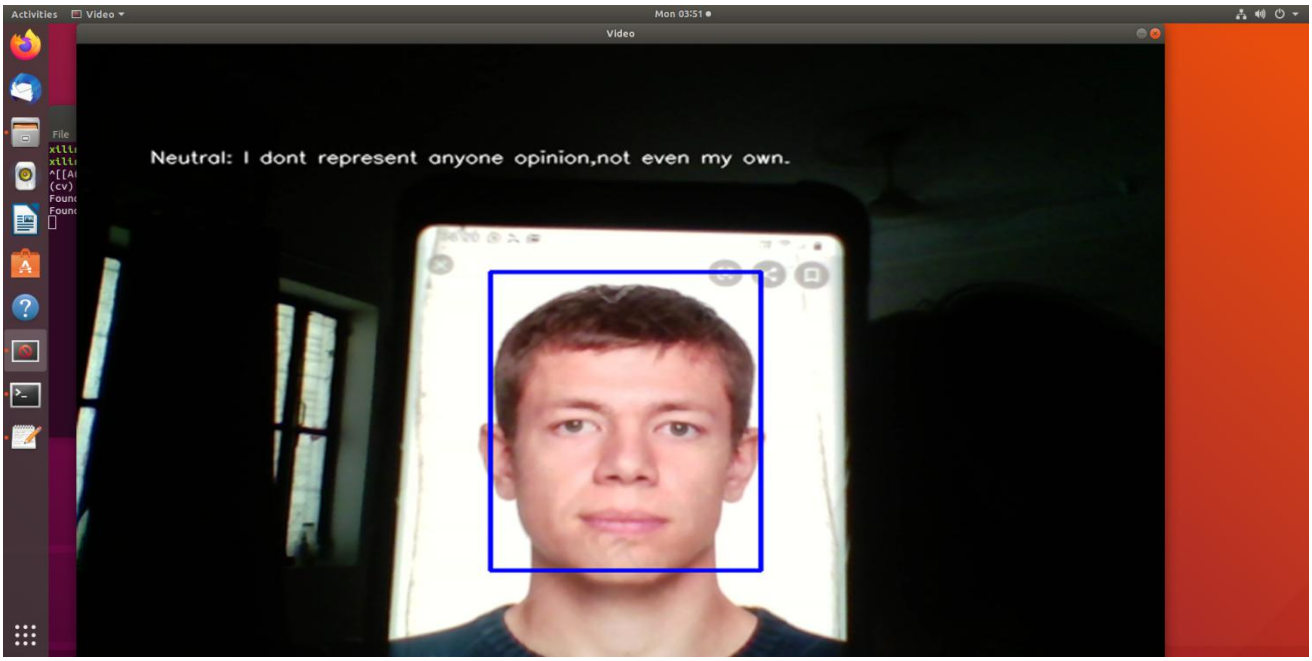


Fig 3.NEUTRAL

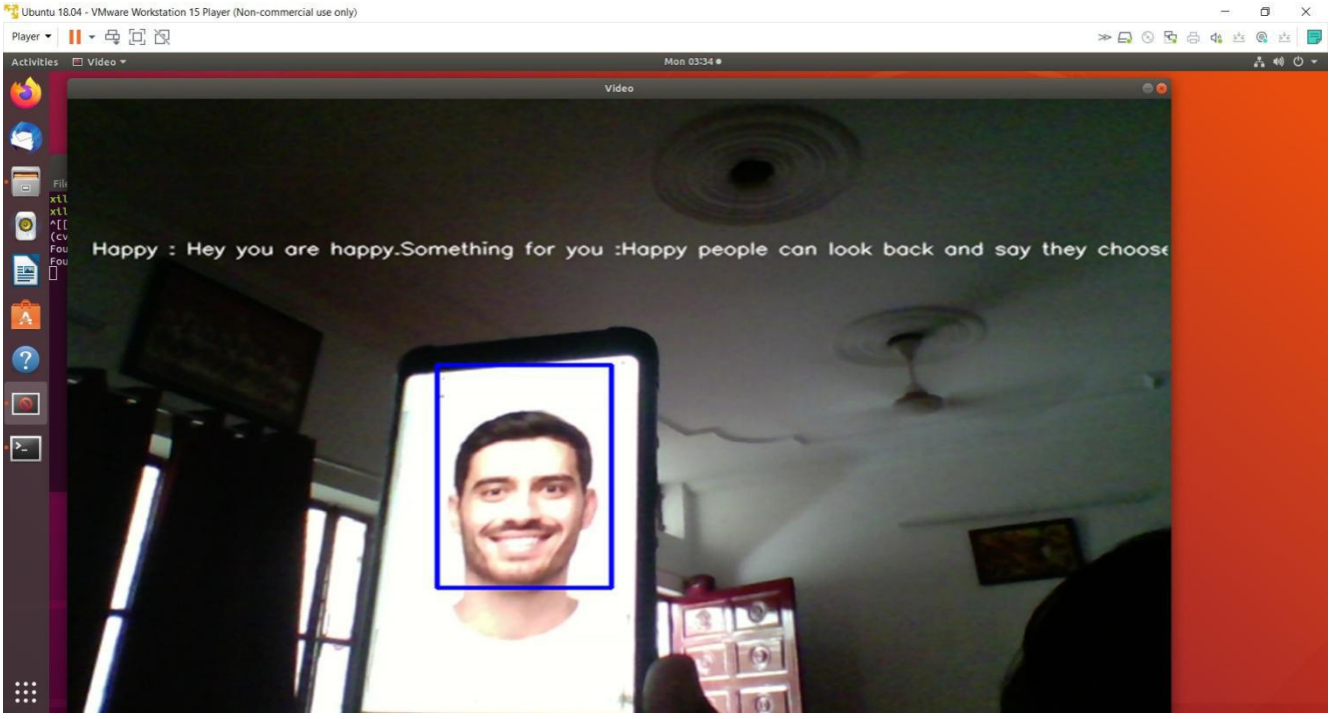


Fig 4. HAPPY

F

CONCLUSION

If the user is feeling sad, then system will automatically fetch a joke from database and send it to the user on the Window terminal. The system is able to make a conversation through chatting application. System is able to send some links, web pages or information by recognizing response from user. It will help to decrease level of stress

on mind. Also support stress management.

REFERENCES

- Mood Detection: Implementing a facial expression recognition system.http://cs229.stanford.edu/proj2009/AgrawalCosgriffMudur.pdf
- Automatic Mood Detection from Acoustic Music Data1 http://esf.ccarh.org/254/254_LiteraturePack1/Emotion1_MoodDetection(Liu).pdf
- https://www.google.com/search?rlz=1C1JZAP_enIN856IN856&sxsrf=ACYBGNTZEXNm52sYtOUUhETDc2JgZu3zvA%3A1581418814506&ei=PolCXou4HtjyrAGw5IaoDw&q=what+do+we+write+under+motivation+in+a+project+report&oq=what+do+we+write+under+motivation+in+a+project+report&gs_l=psy-ab.3...1019971.1034392..1035200...0.1..3.1505.16429.0j7j29j6j3j1j1j1j1.....0....1..gws-wiz.....0i71j0i67j0j0i273j0i131j35i39j0i20i263j0i22i30j33i160j33i22i29i30.oWdCdDvS21M&ved=0ahUKEwiLz7eVrMnnAhVYOSsKHTCyAfUQ4dUDCAs&uact=5
- https://tisonprojects.com/ai-desktop-partner/
- https://www.engpaper.com/cse/artificial-intelligence-ai-desktop- partner.html