IJRASET

International Journal For Research in
Applied Science and Engineering Technology

Song Recommendation Using Facenet Emotion Technology

Abhinay Chaukade[1], Hannah S[2]

[1, 2]Veltech Rangarajan Dr. Sagunthala Institute R&D Institute of Science and Technology, Avadi, Chennai

Abstract: In today's technological revolving era, everyday new types of technologies are being discovered. Various types of technologies, devices, softwares etc. are being introduced daily. According to the experts the facial expression is derived from the internal biological motions of the muscles that lie under the skin. It also deprives the emotional state of the person. But, in this world where communication has become an important thing, the facial expression is the mean of the non-verbal communication, by which we can understand the current behaviour of the person without even uttering single word. For recognition of an individual's mood, the system uses Facial Expression Recognition (FER) concept and depending on the recognized mood, song is played.
This program removes the time-consuming and the hard work of manually searching and playing the songs from internet or any other sources. So, we introduced an application that suggests the songs based on the facial expression or emotions. This application will use the camera feature of any device to read the facial expression of the user and then define a song accordingly. Furthermore, this application will have an additional feature of offline functionality. These are the modules that our application will function on. Our application will imply on various algorithms and packages and application interfaces. We can precisely say our application will use Python package of Tensor Flow (TF) for the facial expression. In this way, the proposed system will contribute towards the healthy emotion building of the humans and will technologically evolve the existing music recommendation system in a new wholesome way.
Index terms: Facial Expression Recognition, Python Package, Tensor Flow, API, BBOX.

I. INTRODUCTION

Globally everyone likes to know about the trending technological developments circulating in the world. Likely, one the technology is our proposed system of song recommendation using face net. There are many existing systems that could recognize facial emotions. On the other hand, there are systems that recommend music. Bringing together, a system or precisely a technology diffused in a program which will recommend music by recognizing the mood of the user from facial emotions is the overall concept described here. Emotion recognition would have larger scope in the near future in fields like robotics for efficient sentimental analysis without the involvement of another human. The enhancement in the accurate prediction of the human emotion will be having a large-scale usability in the near future. The emotion, mood, sentiment, expression that is generated by the user is a biological process but to read them by a machine and to identify and read them, it's a challenge that will be fulfilled up. Lest there is no official music players that offers the feature of playing music via emotion recognition. Also, various applications have the limitations to operate without proper internet connection, but in our application, there is offline availability on which user can operate without internet connection. Thus, all these combines and make our program ready to respond according to user's requirements.

II. RELATED WORKS

A. Literature Survey

This paper [1] states the information about the work upon the senses like visual, auditory, tactile and etc. Here they worked upon the confusion matrix as true positive and true negative. For a perfect system it should obliged the rule of high true positive and very low false positive. Three types of uprooting methods are used here: Generic method, feature template method, structural matching method. The first approach pools all the images and constructs a set of eigen faces that represent all the images from all the views. The other approach uses separate eigen spaces for different views, so that the collection of images taken from each view has its own eigen space. Using a limited set of images (inclusion of all image set angry vs happy or sad vs angry etc, varying views of person), recognition performance as a function of the number of eigen vectors was measured for eigen faces only and for the combined representation

The paper [2] defines about the first approach pools all the images and constructs a set of eigen faces that represent all the images from all the views. The other approach uses separate eigen spaces for different views, so that the collection of images taken from each view has its own eigen space. Using a limited set of images (45 persons, two views per person, with different facial expressions such as neutral vs. crying), performance of recognition as a function measured for eigen faces only and for the combined representation for the number of eigen vectors. Comparison of matching: (a) test views, (b) eigen face matches, (c) eigen feature matches. Earlier processes treated facial recognition as the standard problem of pattern recognition; later processes concentrated more on representation, after realizing the uniqueness (by usage of domain knowledge); more recent methods have been concerned with both representing and recognizing, so that an efficient system with good generalizing capacity could be built up.

This paper [3] discusses the module, that Viola-Jones (VJ) Algorithm creates a detector for detecting the facial images and the Box Bounding technique in which important objects like Nose, Eyes, Mouth are highlighted with a boundary. To detect the faces from the input image, a system object (facial) detector is created with the help of command used by the Viola-Jones algorithm is: detector vision. Cascade Object Detector.

Once the object detector is made, with the help of this given syntax the step method is called: BBOX=step (detector, I) Which outputs BBOX which is M*4 matrix consisting of M bounding boxes having the detected objects. This module consists of the PCA approach in which higher dimensional images are converted into lower dimensional ones, thus retaining the primary image information. The average face i.e., mean is calculated and are then subtracted from training data set images. From the input test image and the eigen faces in the data set, Euclidean Distance is calculated. This was the precisely researched information stated by this paper.

This paper [4] informs about the comparison of other algorithms used in previous systems, the proposed algorithm is proficient enough to battle large pose variations. On the other hand, rarely, some other algorithms detect and locate the faces at the same time. To depict progressively trademark highlights of the specific chose face most noteworthy Eigenvalues of the eigen vector will be picked as the ideal eigen face. Music Feature: Music can be recommended based on available information such as the album and artist. The data set we utilized for preparing the model is Million Song data set. The songs are disintegrated into several music pieces and the mood of the song is recognised.

This paper [5] guides about the minimum and maximum window size is chosen, and for each size a sliding step size is chosen. Then the detection window is moved across the image as follows: Set the minimum window size, and sliding step corresponding to that size. Each face recognition filter (from the set of N filters) contains a set of cascade-connected classifiers. By using deep learning and neural networks the emotion is detected in an accurate manner. CNN2 is used in the proposed system in order to achieve cost efficiency. In the area of face location, the framework yields identification rates practically identical to the best past frameworks. The most recent logical discoveries show that feelings assume a fundamental part in basic leadership, observation, learning, and that's only the tip of the iceberg that is, they impact the very components of reasonable reasoning.

III. PROPOSED WORK

A. Existing System

The existing systems that are being used in day today's life are manual searching (i.e. typing or voice command) of songs on the platform or the utmost requirement of the prompt data connection for the operability and mainly the Artificial Intelligence recognizing the face of its owner to unlock things are some of the existing systems that we do use daily.

B. Proposed System

The system proposed by our team is to build an application that will be suggesting the songs based on the facial expressions or emotions, the accurate prediction of song based on emotion and finally the application will be able to operate without prompted data connection, which basically any app doesn't provide. These are some of the top-notch features that our program will provide to end-users. The program will keep captured images in records for track. This is the proposed system by our team.

C. Hardware Specification

1) RAM 3 gb ROM 32 gb [minimum].
2) 1.6 Ghz Quad-core processor [minimum].
3) 8-megapixel front camera [minimum].
4) Device must have location services [required].
5) Device at-least have HSPA+ cellular network [required].
6) Dedicated speaker [required].

IV. IMPLEMENTATION AND ANALYSIS

A. Description

Facial recognition and emotion sensing begin from what the user is displaying or querying the application about. Typically to rationalize and simplify this scenario we have divided the general proforma in three stages: initiation stage, processing stage and finally the output stage.

B. Initiation Stage

In this stage, various procedures and actions like starting of the application or the program, then asking for the permission, emotion recognition of the user and find face and provide the song based on emotion and lastly the offline accessibility of the app (without data connection). These are all first stage input actions that if they are selected, they'll navigate to the other stage.

C. Processing Stage

In this stage, the processing of the data based on the user input is performed, so the emotion recognition and song recommendation totally relies on the data libraries of the application (i.e. mood defines what song will be suggested), the test images get stored in the database for the future use and track records.

D. Output Stage

In this stage, after applying algorithm and analysis techniques of the processing stage, storage of content/data, the final output is displayed for the user queries and requirements (i.e. songs based on face emotion, offline song if no data connection found, and if the user closes the app; terminate all the processes. The further description about the module can be found in the figure-1 given below.
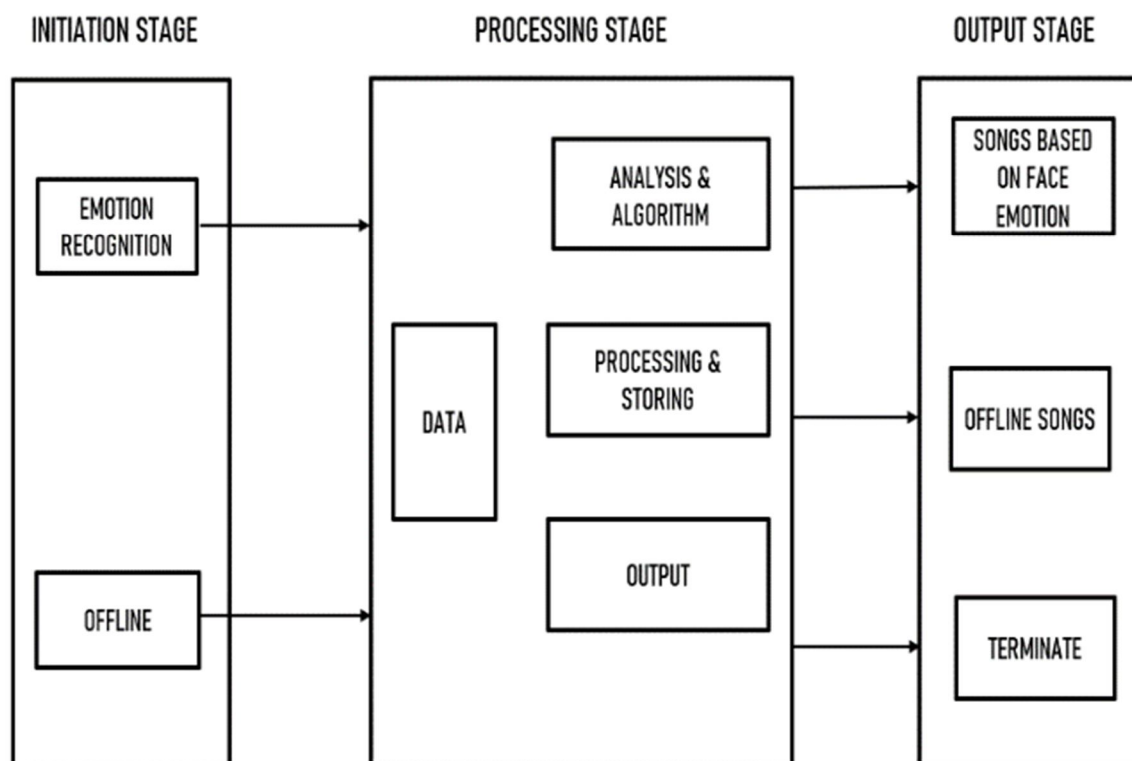

Fig 1: General Architecture of Proposed System

The designing phase begins with the designing of the modules along with considering all the aspects and following the project criteria.

E. Design Description

Well, the data of data will take place in accordance with the diagram. Firstly, the main application/program will be launched by the user and the launch activity will get started. Then the launch activity will lead to the authentication of the packages and libraries that will be used to run this program. After the successful authentication of libraries and permission, then the program will drift to the graphical verification will be executed. Then the opening of camera takes place scanning of face is performed. After successful scanning, the emotion identified is shown. Finally, the song will be played based on the emotion identified. This is the final output.
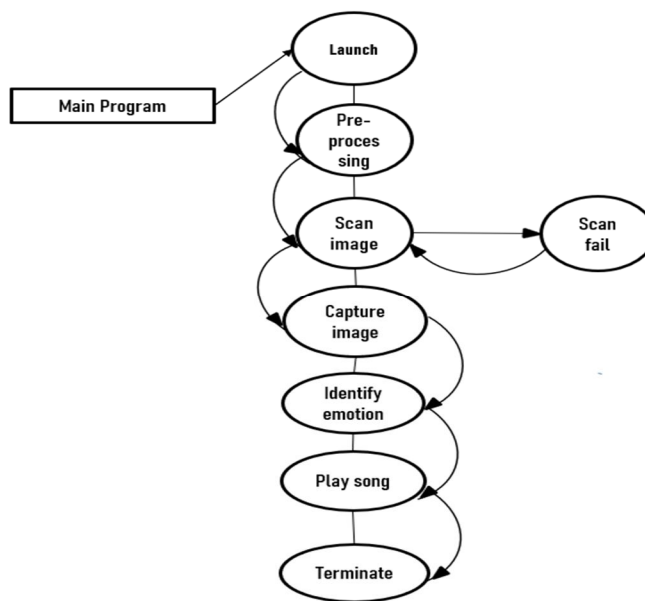

Fig 2 Data Flow Diagram of Proposed System

F. Modelling Description

The UML diagram shows when the user performs some action, user will definitely receive some response for that. So, when the user enters in the program/application, everything will be authenticated and then he will be granted access. Similarly, the action of face recognition will analyse and process the binary image and then will find the result stores it and finally will display the output. And so, on it goes for the offline accessibility.
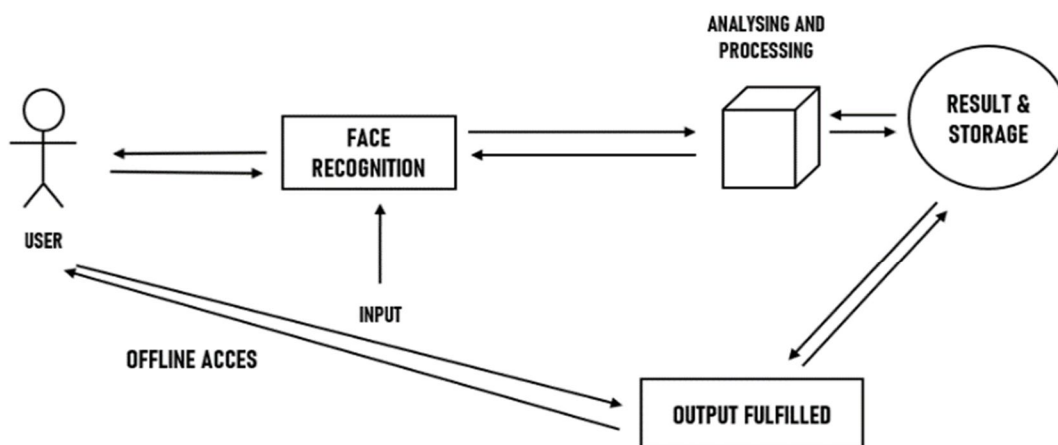

Fig 3: UML Diagram of Proposed System

G. *Collaborative Description*

The description of the use-case diagram, sequence diagram and collaboration diagram are same as the before defined diagram. The concepts are thoroughly inter-connected with each other.
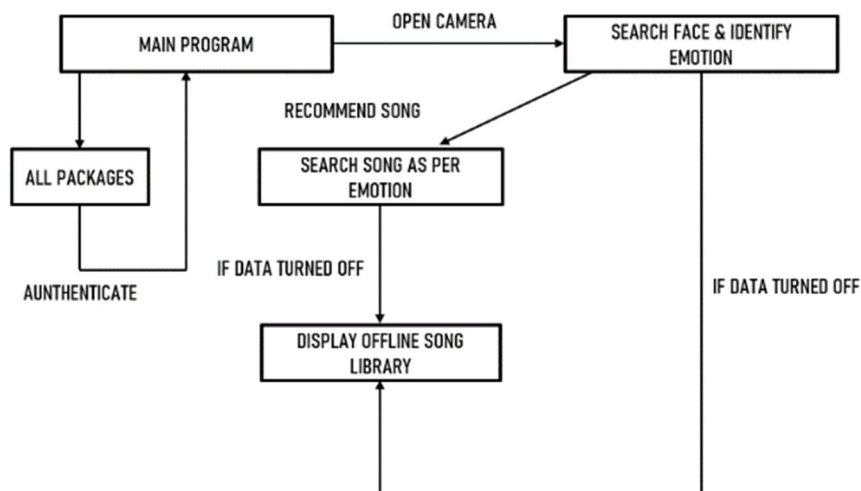

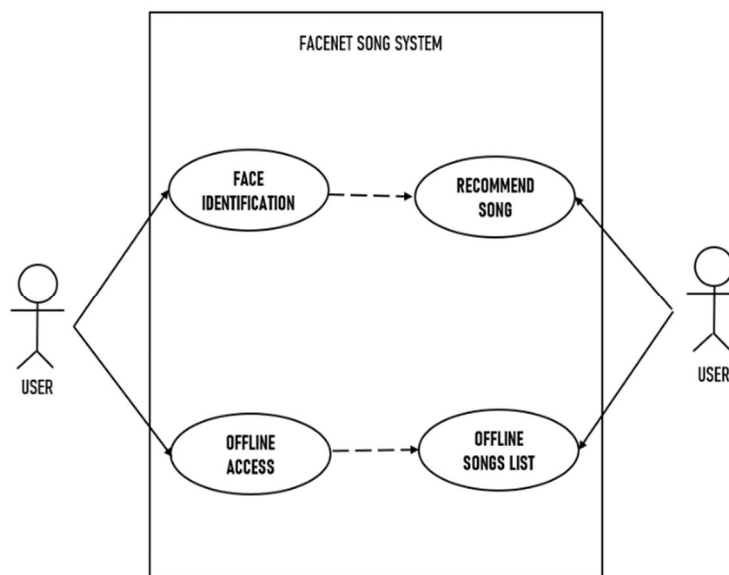Fig 4: Class Diagram of Class Diagram

1) *Use Case Diagram*

Fig 5: Use Case Diagram of Proposed System
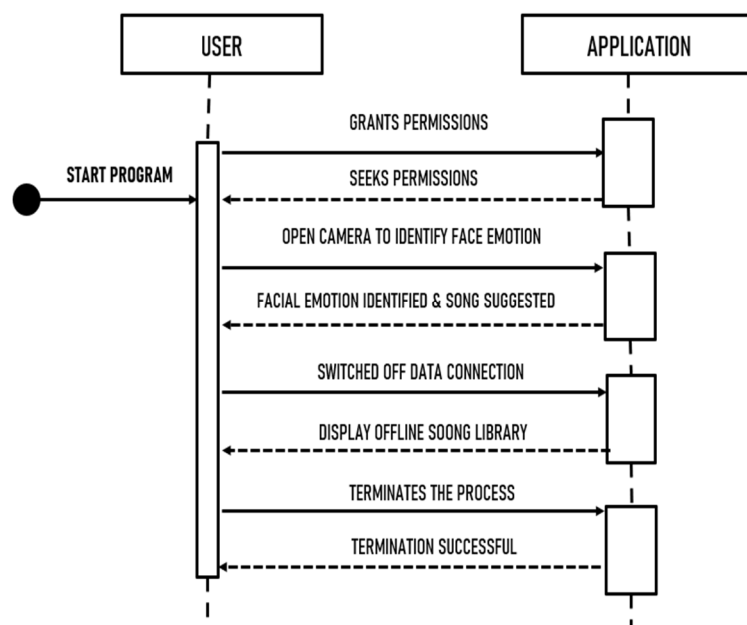
2) Sequence Diagram

Fig 6: Sequence Diagram of Proposed Diagram

V. METHODOLOGY

The methodology applied in this system is that the training of playing the song according to the images (with different emotion) are performed and then we obtain the accuracy results of the process by this confidence percentage we get to know how accurate our program is. The training process of the program is done by the TensorFlow package. After this, when the training gets completed, we need to install some important packages. The various algorithms utilised are the Viola-Jones algorithm, Principal Component Analysis, Facial Expression Recognition algorithm, Haarcascade etc.

While actually performing the process, we get the camera access via webcam in which we have to provide our face from front view and then the program will make a frame surrounding the face displayed with emotion tagged above (i.e. Happy, Sad, Angry, Neutral etc). After correctly identifying the emotion of the face displayed, then the program will search for the best suited song according to the current mood. Along with the key shortcut to pause, play, exit the song etc. After the process is terminated the image displayed in the camera for facial detection, will be saved as "test.jpg" in the project destination.

The module description are as follows:

A. Provide Input

Since our application is built on the platform "PyCharm IDE", the coding will be specifically done on the platform in the "Python" language, the first module of our program was to insert the various emotion distincted images of human faces.

B. Adding Emotion Recognition

The second module of the application after successfully inserting human emotion images in the program, now we will be adding the feature of facial/ emotion recognition. This will be using the coding "python" to be up to the required constraints.

C. Adding Songs in Library

The third module of our program is to add the various songs of different emotional and sentimental feels. The happy songs, the sad songs, the songs that portrays anger, and etc. We'll make a library of songs and add all songs to it.

D. Training the data set using TENSOR FLOW

The fourth module of our program is to execute the training of the data set that be mandatorily requiring the "TENSOR FLOW" package Now the training data set command will be executed in command prompt and training will be executed.

E. Offline Accessibility feature

The final module of the application is to add the feature of offline accessibility that will allow user to access offline libraries of the song without prompting data connection. This feature will be the last module of the application. Hence, these were the description of the modules that will be used in the application so make the end result possible. These all module gathers up and then makes up the final end application that will be tested on different devices to check the efficiency and then released. Thus, the introduction of module is finished. The input and output flow of the system is explained in the diagram given in the figure given below.

F. Input Design

The input design will be the input actions that will be given by the user requirements according to its needs. Firstly, the user's device compatibility will be authenticated then if the authentication was successful then the mood/emotion recognition will get open. Another, input will be the offline accessibility

G. Output Design

The output will be based on user's camera quality is low to scan or identify the emotion, then process will get failed and again start back from starting. If the input of emotion is fully correct and scanning gets completed, then output will be the song playing based on that emotion which was identified earlier. And if program identifies the input as no internet connection, then output will be songs in offline library. These all will be displayed as output which will be totally depending on the user's input feed.
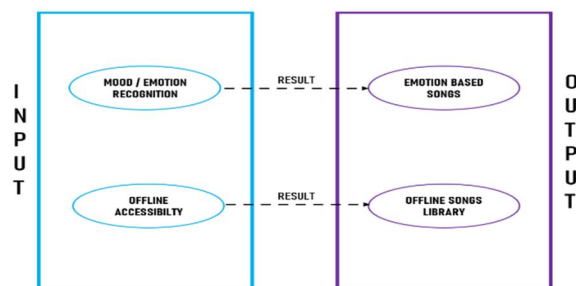

Fig 7: Sequence Diagram

VI. OUTPUTS AND RESULTS

The efficiency of the proposed system was up to the expectation, the accuracy, user interface, database functionality etc are working and yielding good results. So, in that we say the efficiency of the proposed system proved good.

REFERENCES

[1] Prateek Sharma, "Multimedia Recommender System Using Facial Expression Recognition" University of Galgotias, International Journal of Engineering Research And Technology, Vol. 9 Issue 05, May-2020, pp. 674-676

[2] Samuvel, D. J., Perumal, B., y Elangovan, M. (2020). Music recommendation system based on facial emotion recognition. 3C Tecnolog´ıa. Glosas de innovacion aplicadas a la pyme. Editi ´on´ Especial, March 2020, pp. 261-271.

[3] Pratik Gala, Raj Shah, Vineet Shah, Yash Shah, Mrs. Sarika Rane, 2018, International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056

[4] Prof. Jayshree Jha, Akshay Mangaonkar, Deep Mistry, Nipun Jambaulikar, Prathamesh Kolhatkar, "Facial Expression Based Music Player", Vol. 4, Issue 10, October 2015, pp. 331-333

[5] W.Zhao, R. Chellappa, P. J. Phllips, A Rosenfeld, "Face Recognition: A Literature Survey" University of Maryland, Vol. 35, No. 4, December 2003, pp. 399–458

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)