IJREAT International Journal of Research in Engineering & Advanced Technology, Volume 8, Issue 5, October - November, 2020
ISSN: 2320 – 8791 (Impact Factor: 2.317)
www.ijreat.org

Classifier for Music Using Facial Emotion Recognition (FER)

[1]Rahul Sailesh Wadhwa Sir , [2]V Binu Aldrin Sir, [3]Mrs Mayuri KP
[1,2] MVIT, Bangalore,
[3]Asst. Professor, Dept. of CSE, Sir Mvit-Bangalore

Abstract **- The human face is an important part of an individual's body and it especially plays an important role in knowing an individual's mood. Extracting the required input from the human face can now be done directly using a camera. This input can then be used in many ways. One of the applications of this input can be for extracting the information to deduce the mood of an individual. This data can then be used to get a list of songs that comply with the mood derived from the input provided earlier. This eliminates the time-consuming and tedious task of manually Segregating or grouping songs into different lists and helps in generating an appropriate playlist based on an individual's emotional features. Various algorithms have been developed and proposed for automating the playlist generation process. Facial Expression Based Music Player aims at scanning and interpreting the data and accordingly creating a playlist based the parameters provided. The scanning and interpreting includes audio feature extraction and classification to get a list of songs belonging to a similar genre or to get a list of similar sounding songs. Human emotions are meant for mutual understanding and sharing feelings and intentions. The emotions are manifested in verbal and facial expressions. Face detection has been around for ages. Taking a step forward, human emotion displayed by face and felt by brain, captured in either video, electric signal (EEG) or image form can be approximated. Human emotion detection is the need of the hour so that modern artificial intelligent systems can emulate and gauge reactions from face. This can be helpful to make informed decisions be it regarding identification of intent, promotion of offers or security related threats. Recognizing emotions from images or video is a trivial task for human eye, but proves to be very challenging for machines and requires many image processing techniques for feature extraction. Several machine learning algorithms are suitable for this job. Any detection or recognition by machine learning requires training algorithm and then testing them on a suitable dataset. People of today's world are very indecisive even when it comes to the simplest of things like music. This project's purpose is to design a classifier that could be used to eliminate this confusion exhibited by people. This project is about building a music recommendation system for users. Such a system can not only be used to brighten up one's mood on a rainy weekend; especially in hospitals, other medical clinics, or public locations such as restaurants, this classifier could be used to spread positive mood among people. The project has to be implemented in 2 phases-facial emotion recognition and then suggestion of music based on the detected emotion in the previous phase. The emotional reaction to music is different for every person, so analysing it will not likely yield perfect results. The method used then is to decide upon certain base songs that very closely embody a certain mood, and to match songs to these specific categories. We can also study the efficacy of music stimuli to modify (reduce rather than induce) naturally occurring negative mood states (like sad, depressed, etc.) and to gradually bring about a change in this negative mood state turning it into a positive one. It could also simply be used just to prevent certain songs from having a negative impact on a person's mood.**

I. Introduction

Human emotion detection is implemented in many areas requiring additional security or information

about the person. It can be seen as a second step to face detection where we may be required to set up a second layer of security, where along with the face, the emotion is also detected. This can be useful to verify that the person standing in front of the camera is not just a 2-dimensional representation [1].

Music is present in all cultures, providing a construct through which mood can be expressed. The ability of music to evoke mood makes it is a potentially powerful resource for mood modulation. Human emotions can be classified as: fear, contempt, disgust, anger, surprise, sad, happy, and neutral. These emotions are very subtle. Facial muscle contortions are very minimal and detecting these differences can be very challenging as even a small difference results in different expressions [2].

Recognition of facial expressions is used to identify the basic human emotions. Facial expressions give important rules about emotions. Computer systems based on affective interaction could play an important role in the next generation of computer vision systems. Face emotion can be used in areas of security, entertainment and human machine interface (HMI). A human can express his/her emotion through lip and eye. Generally, people have a large number of songs in their database or playlists. Thus to avoid trouble of selecting a song, most people will just randomly select a song from their playlist and some of the songs may not be appropriate for the current mood of the user and it may disappoint the user. As a result, some of the songs are not matching to the user's current emotion. Moreover, there is no commonly used application which is able to play songs based on the current emotions of the user. Music plays a very important role in enhancing an individual's life as it is an important medium of entertainment for music lovers and listeners and sometimes even imparts a therapeutic approach. In today's world, with ever increasing advancements in the field of multimedia and technology, various music players have been developed with features like fast forward, reverse, variable playback Although these features satisfy the user's basic requirements, yet the user has to face the task of manually browsing through the playlist of songs and select songs based on his current mood and behaviour [3].

The main objective of this project is to design an efficient and accurate classifier that would generate a playlist based on current emotional state and behaviour of the user. Face detection and facial feature extraction from image is the first step in emotion based music player. For the face detection to work effectively, we need to provide an input feed [3]. We have used an algorithm that is used for face detection and facial feature extraction. We have generated landmarks points for facial features. This is done by means of a pre trained HAAR cascade classifier. The next step is the classification of emotion for which we have used a FisherFace classifier. The generated landmarks points are provided to the FisherFace classifier for training purpose. The emotion classified by classifier is then passed to music player and accordingly music will be played.

II. Literature Review

Facial expressions are one of the most important nonverbal channels for expressing internal emotions and intentions. Traditionally, algorithms for automated facial expression recognition consist of three main modules, viz. registration, feature extraction, and classification. Detailed survey of different approaches in each of these steps can be found in [4]. Conventional algorithms for affective computing from faces use engineered features such as blockwise Histogram [5], dictionary approaches [6], facial landmarks [7,8,9,10]. Each of these methods is specific to the method of implementation and in certain cases a combination of methods has been utilised as well. Since the majority of these features are hand-crafted for their specific application of recognition, they often lack required generalizability in cases where there is high variation in lighting, views, resolution, subjects' ethnicity, etc.

One of the methods for recognition of emotions could be by the use of facial landmarks (mainly the eyes, mouth and eyebrows) and then mapping these facial landmarks to certain cues for recognising the different emotions as it has been done in [7,8,9,10]. Once these facial landmarks have been extracted, different methods use different set of machine learning algorithms to map the landmarks to a certain expression. The different algorithms that can be defined are shown in [10]. Out of all these it can be observed that the KNN and Random Forest algorithms gave the highest accuracies.

For the extraction of the facial landmarks we can make use of various methods such as the use of OpenCV as in [9], OpenCV and then Dlib as in [7], face alignment method and the typical active appearance model (AAM) which has been utilised in [8], or even the LuxandFaceSDK as in [10]. Once these landmarks are extracted Manhattan/Euclidean distance is used in different machine learning algorithms (KNN, SVM, graph-based methods and

Random Forest) for classification of emotion. The classification was done either using logistic regression or direct classification.

Another approach to determining the emotion would be to make use of dictionaries as has been done in [6]. In this approach the author makes use of block-wise histograms to create a feature dictionary. The advantage of this method was that it was able to reduce the feature dimensions while inheriting the advantages of domain specific languages (DSLs).

For a song classifier the method of implementation was the use of an excel spreadsheet which contained a classified set of songs that were sorted on the base of emotions. These songs were stored as .mp3 files in a folder which contained more sub-folders classified on the basis of emotions. Finally, a simple call was made to the .mp3 file so it could be played in the music player.

III. Emotional Taxonomy

Emotion classification, the means by which one may distinguish or contrast one emotion from another, is a contested issue in emotion research and in affective science. Researchers have approached the classification of emotions from one of two fundamental viewpoints:

1. that emotions are discrete and fundamentally different constructs
2. that emotions can be characterized on a dimensional basis in groupings [11]

In discrete emotion theory, all humans are thought to have an innate set of basic emotions that are cross-culturally recognizable. These basic emotions are described as "discrete" because they are believed to be distinguishable by an individual's facial expression and biological processes.[12] Theorists have conducted studies to determine which emotions are basic. A popular example is Paul Ekman and his colleagues' cross-cultural study of 1992, in which they concluded that the six basic emotions are anger, disgust, fear, happiness, sadness, and surprise. Ekman explains that there are particular characteristics attached to each of these emotions, allowing them to be expressed in varying degrees. Each emotion acts as a discrete category rather than an individual emotional state.[13]

Fig 1. Different Expressions of Human

Fig 2. Expression of Human Emotions

IV. Dataset

To simulate our proposed model Extended Cohn-Kanade Dataset (CK+) has been used. It contains 593 pictures for 123 subjects of the following emotions (neutral, sadness, surprise, happiness, fear, anger, contempt and disgust) which are mostly grayscale. The resolution of each image was 640*490. Each emotion had a facial expression label and FACS. From the readme of the dataset, the encoding is: {0=neutral, 1=anger, 2=contempt, 3=disgust, 4=fear, 5=happy, 6=sadness, 7=surprise}.

V. Proposed Work

The project has 2 main phases that have to be constructed and integrated. Each one deals with the one part of the problem.

Phase 1 consists of the following:

- Pre-Processing of Dataset (CK+)

- Collection of Image Sample
- Face Detection
- Emotion Recognition

Phase 2 consists of the following:

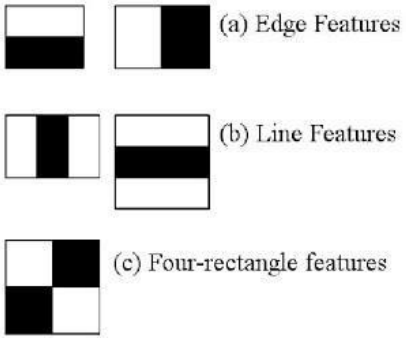
- Formulation of the Dataset (Music)
- Suggestion of Music based on the recognized emotion.

A. PHASE 1

1. Pre-Processing of Dataset
   The CK+ dataset has to be unzipped and all the images for the various emotions depicted by the users has to be cleaned and any duplicate images for the same emotion have to be removed in-order to make the dataset more easily manageable. This pre-processing involves segregating the different images into a set of folders that represent the different emotions which is either done manually or by making use of an existing classifier. The haar cascade XML file is formed in this stage.

2. Collection of Image sample
   The system needs image input that comprises of the different emotions that can be used to depict the different moods of the user. For that, the system has a predefined set of images that are previously classified by the CK+ dataset. So, the emotion that the system needs to classify, is done using the live input given by the user. On receiving the image sample, the system will pre-process it and store it into the various classified folders.

3. Face Detection
   We will use a pre-trained HAAR classifier supplied with OpenCV. 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. Each feature is a single value obtained by subtracting sum of pixels under white rectangle from sum of pixels under black rectangle.



First, we need to load the required XML classifiers. Then load our input image (or video) in grayscale mode. We make use of 'haarcascade_frontalface_default.xml' in order to obtain a box around the face detected. Now we find the faces in the image. If faces are found, it returns the positions of detected faces as Rect(x,y,w,h). Once we get these

$$S_b = \Sigma_{j=1}^{C} (\bar{N}_j - N)(\bar{N}_j - N)^T$$

$$S_w = \Sigma_{j=1}^{C} \Sigma_{i=1}^{Nj} (x_i^j - \bar{N}_j)(x_i^j - \bar{N}_j)^T$$

$$W_{Frisher-Descriminat} = \arg\max \frac{|W^T S_b W|}{|W^T S_w W|}$$

locations, we can create a ROI for the face.

4. Emotion Detection
   To detect the actual emotion on your face we use an Individual model rather than a generalized model. Models trained on a single individual work much better when used on the same individual, often because in that case there is less variance between the data (here :facial features). If we minimise the variance by keeping the face the same, most of the detected differences will be due to the fact that a different emotion is expressed. You need to collect varied images of yourself to make this robust. We make use of Fisherfaces algorithm for emotion classification. Fisherfaces is a supervised classification method used for face recognition. It is used to classify samples of unknown classes based on training samples with labelled classes. The goal of Fisherfaces is to maximize between-class (across users) variance and minimize within-class (within user) variance. Here, C class is

the face with mean of class j denoted by N j – and the ith image in class j denoted as j xi . These scatter matrix are sb and sw respectively and are calculated as follows:

B. PHASE 2

1. Formulation of Music Dataset
   The music dataset has to be created and hosted on the system. This includes sorting the various songs that are .mp4 files and storing them in folders that will be assigned with the emotion labels. There is no limit on the number of songs that can be stored on the system. This stage involves getting together the entire music dataset and sorting them according to emotion in the folders as well as in a .csv file.

2. Suggestion of Music based on Recognized Emotion
   Finally after the emotion has been recognized, a playlist has to be generated according to the emotion recognized by the classifier. This is done by a simple retrieval of data from the .csv file which already consists of the music that has been categorized into the various emotional states defined. Once this is done the user can select a song from the list displayed and the music can then be played easily. The user can also add or delete songs to the list as per his/her whims and fancies.

VI. Results and Analysis

Unit Testing- This is the first level of testing, where different components are tested against the requirement specification for the individual components. Unit testing is essential for verification of the code produced during the coding phase and hence the goal is to test internal logic of those components.

System Testing- System testing is performed on the entire system in the context of functional requirements specification and system requirement specifications. System testing tests not only the design, but also the behaviour and the expectations. It is also to test up to and beyond the bounds defined in the software requirement specification. System testing for our platform is accomplished by the following test cases. These test cases show how each and every component interacts with each other as a dynamic system.

Fig (a), (b), (c), (d), (e) are the results of running the implementation of our proposed work.

Fig (a) contains an accurate description of the different trial and errors we got using our source code.

Our music player has been designed to recognize only angry, happy, sad and the neutral expressions. Using only these categories we got 77.2% accuracy. This means that almost 4 out of 5 times it will play a song fitting to your emotional state.

TEST CASE ID	PAGE	TEST CASE	EXPECTED RESULT	ACTUAL RESULT	STATUS
TC01	Home screen	Face Detection	Face is detected	Face is detected	PASS
TC02	Home screen	Face not present	Show error	Show error	PASS
TC03	Home screen	Play music without selection of music.	Show error	Show error	PASS
TC04	Home screen	Happy Emotion Recognition	Emotion Recognized	Emotion Recognized	PASS
TC05	Home screen	Neutral Emotion Recognition	Emotion Recognized	Emotion Recognized	PASS
TC06	Home screen	Sad Emotion Recognition	Emotion Recognized	Emotion Recognized	PASS
TC07	Home screen	Anger Emotion Recognition	Emotion Recognized	Emotion Recognized	PASS
TC08	Playlist screen	Playlist generation based on emotion	Playlist Generation for respective emotion detected	Playlist Generation for respective emotion detected	PASS
TC09	Music screen	Play selected music	Play the selected music	Play the selected music	PASS
TC10	Home screen	Detect new face and emotion	New emotion detected	New emotion detected	PASS

Fig a. Neutral Emotion Identification Result

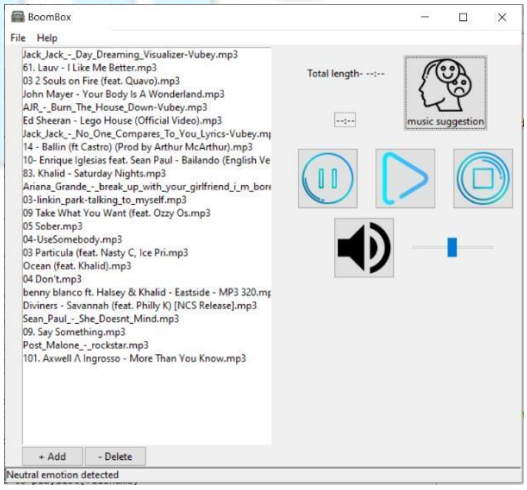


Fig b. Neutral Emotion Identification Result

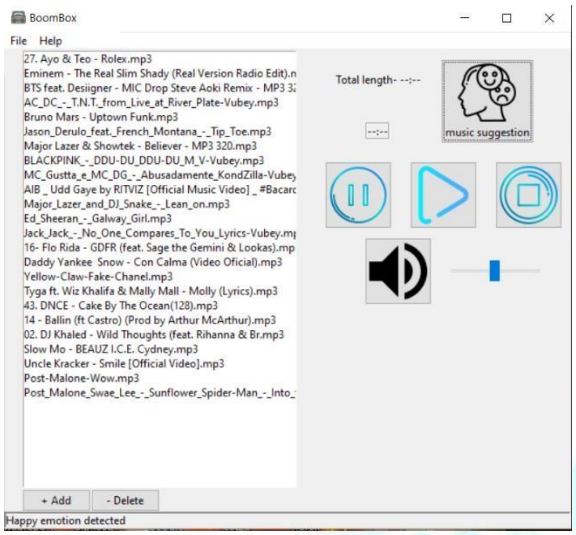


Fig c. Happy Emotion Identification Result

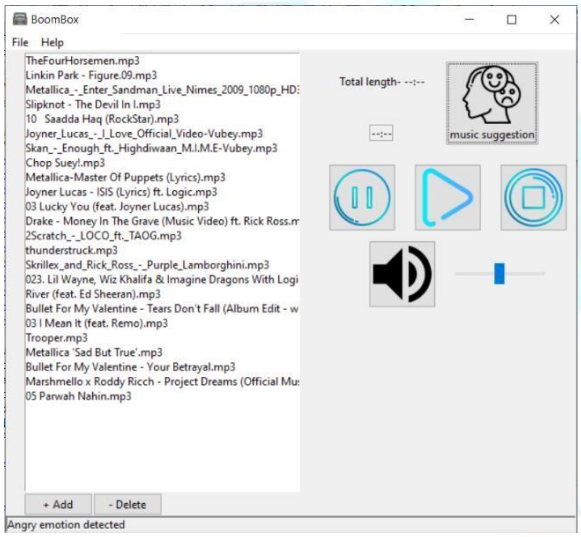


Fig d. Angry Emotion Identification Result

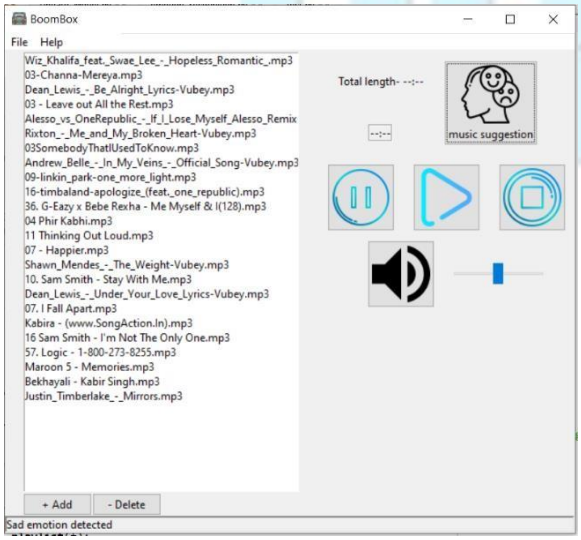


Fig e. Sad Emotion Identification Result

VII. Future Work

The system can be improved in the following ways:

- Enable the classifier to identify a wide range of emotions.
- Personalized classification of songs based on the preference of the user.
- Can detect when a person is sleepy during driving.
- Addition of an Audio detection component to increase the accuracy of the classifier by sensing the tone of the voice. Algorithms like logistic regression, linear discriminant analysis and random forest classifier can also be fine-tuned to achieve good accuracy and results.
- Integration of the classifier into a web/mobile application.

VIII. Conclusion

Our implementation can roughly be divided into 2 phases:

Phase 1 consists of the following:

- Pre-Processing of Dataset (CK+)
- Collection of Image Sample
- Face Detection
- Emotion Recognition

Phase 2 consists of the following:

- Formulation of the Dataset (Music)
- Suggestion of Music based on the recognized emotion.

Feature extraction is a very important part of the experiment. Our objective is to design a model that can identify your emotion and then implement a real time model that can be used to suggest music by utilising the categorized emotion of a given feed. We are determining the emotion of a particular person

from a webcam, the webcam should be able to detect all the faces accurately. We have implemented a fully functioning system that can suggest music based on a picture taken using the webcam of a laptop, uses this picture as input to the system and finally suggests music to the user based on the emotion portrayed by him in the picture taken.

IX. References

1. W. Swinkels, L. Claesen, F. Xiao and H. Shen, "SVM point-based real-time emotion detection", 2017 IEEE Conference on Dependable and Secure Computing, Taipei, 2017.

2. 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

3. Prof. Vijaykumar R. Ghule, Abhijeet B. Benke, Shubham S. Jadhav, Swapnil A. Joshi, "Emotion Based Music Player Using Facial Recognition", IEEE Vol. 5, Issue 2, February 2017.

4. E. Sariyanidi, H. Gunes, and A. Cavallaro. Automatic analysis of facial affect: A survey of registration, representation, and recognition. IEEE transactions on pattern analysis and machine intelligence, 37(6):1113–1133, 2015.

5. D. Yang, Abeer Alsadoon, P.W.C. Prasad, A. K. Singh, A. Elchouemi, "An Emotion Recognition Model Based on Facial Recognition in Virtual Learning Environment", Elsevier, 2018

6. Keyu Yan, Wenming Zheng, Zhen Cui, Yuan Zonga, Tong Zhang, Chuangao Tang, "An extended dictionary representation approach with deep subspace learning for facial expression recognition", Elsevier.com,2018.

7. Nitisha Raut, "Facial Emotion Recognition Using Machine Learning", SJSU scholar works, 2018.

8. Nianyin Zeng, Hong Zhang, Baoye Song, Weibo Liu, Yurong Li, Abdullah M. Dobaie, "Facial expression recognition via learning deep sparse autoencoders", Elsevier.com,2017.

9. Behzad Hasani and Mohammad H. Mahoor, "Facial Expression Recognition Using Enhanced Deep 3D Convolutional Neural Networks", IEEE conference on CVPR,2017.

10. Ratna Astuti Nugrahaeni, Kusprasapta Mutijarsa, "Comparative Analysis of Machine Learning KNN, SVM, and Random Forests Algorithm for Facial Expression Classification", ieee.org,2016.

11. https://en.wikipedia.org/wiki/Emotion_classific a tion

12. Colombetti, Giovanna (August 2009). "From affect programs to dynamical discrete emotions". Philosophical Psychology, CiteSeerX 10.1.1.728.9666.doi:10.1080/09515080903153600.

13. Ekman, Paul (1992). "An Argument for Basic Emotions". Cognition and Emotion. CiteSeerX 10.1.1.454.1984