Sai Venkatesh Aravapalli, 2021, 9:3
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

Emotion Based Music Player Using Support Vector Machine (SVM)

Sai Venkatesh Aravapalli
Department of Computer Science and Engineering,
Vellore Institute of Technology,
Vellore, Tamil Nadu, India.
aravapalli100@gmail.com.

Abstract- Music can have a tremendous impact on a person's psychological behavior. The average Indian enjoys music for four hours a day. People love to listen to music to suit their mood and interest. This project helps us to play and listen to music by capturing facial expressions while reflecting on our mood. Facial expression is an unprepared meeting of communication. Computer vision helps the computer to transmit digital images or videos. In this system, computer vision components play an important role in determining the user's emotions by reading their facial expressions. When the user's emotion is detected, the system opens the playlist for the recorded expression. This saves the user a lot of time by selecting playlist songs and reducing the time taken to play. It monitors user details such as how often a song is played, plays songs based on their interests, and resets the playlist at all times. It tells the user that he is not using the song and suggests to edit or delete it.

Keywords:- Facial Recognition, Music Player, Support Vector Machine.

I. INTRODUCTION

Music is a major part of the player in our daily lives. Music lovers can think of life as a mess when they can't find the right track. This is where the Emotion Based Music Player comes into play. It detects user's facial expressions and plays music according to their mood. The first step is to know the user's facial expressions. The input image should be clear with no blurring or tilt. The Viola-Jones algorithm is used for face recognition and expression extraction.

Landscape points are created for facial features. Classification of emotion is done using multiclass SVM classification. The landscape points generated for SVMs are provided for training purpose. Spirit classified by SVM is then sent to the music player and the music is played accordingly.

Using traditional music players, the client must be physically confused by his playlist and select tunes that reduce his mind-set and lively feel. The mistake was put forward and regularly encountered the problem of tunes coming to the right place.

The presentation of Audio Emotion Recognition (AER) and Music Information Retrieval (MIR) provided a ritual framework with a component that naturally transcends playlists in light of a variety of emotions. While the AER sets up a sound flag in light of some sound headlines under different classes of emotions, MIR is an area that relies on the investigation of important data, distinguishing sound headlines from the sound flag.

Although AER and MIR have enhanced the capabilities of traditional music players by eliminating the need for manual isolation of playlists and descriptions in view of customer sentiments, these frameworks have not been used as powerful tools for music players. Bound up with human emotions.

Because current calculations are less accurate because they yield singular results and create simple memory overheads of simple increments, data recovery calculations are less effective. In the shortest time they cannot collect huge data from

sound flags. The current sound feel receipt uses a computational displacement model combined with the customer's impression. In addition, the best plans in the classroom will be rejected, enough to promote edited playlists by devoting human emotions endorsed by the image of the face, without impairing the extra assets. Current architectures use additional tools (such 6 | P a g e as the EEG framework and sensors) or human speech.

This paper then proposes an activity approach Limiting the disadvantages and weaknesses of existing innovation. The hidden goal of this paper is to create an accurate calculation that generates a set of tunes from a subscriber's playlist to suit the customers 'exciting state. Computed planning, less computation time, reduces stockpiling and increased costs in the use of additional equipment. It commands a photo of the face in 4 different exterior displays. Rarely, anger, antagonism and happiness.

II. RELATED WORK

The proposed calculation in this algorithm includes an emotion music suggestion framework that gives the age of an altered playlist in agreement to the user's passionate state.

The proposed strategy includes following modules:

- Input Image
- Testing Image
- Face Detection
- Landmark Point Extraction
- Training Data
- Music Player

1. Input Image:
According to architecture catching picture is the main undertaking to be performed. We are going to catch picture of client by utilizing webcam. There are sure conditions while catching picture, for example, client ought to be close to camera in the event of various clients and furthermore face ought not be tilted.

2. Training Image:
Training picture is given by picture database. Database utilized is JAFFE.

3. Face Detection:
Just face is recognized out of entire picture by playing out specific calculations. We are utilizing Viola-Jones algorithm for face location.

4. Landmark Point Extraction:
In this module 68 Landmark point are recognized and gave to SVM to preparing and testing.

5. Training:
The Data for preparing SVM is put away in record and relating names are additionally put away. The preparation information comprise of 68 milestone focuses separated from pictures of the JAFFE database.

6. Music Player:
This is actual music player where all the songs arranged based on user's emotions are shown to user and user can choose the song according to his will.

III. LITERATURE SURVEY

Various techniques have been proposed to classify human emotions. Face detection is the basic key to this project. The face detected should for be blur or tilted for it to be effective. Many algorithms have been proposed for the face detection to fulfil the above requirements.

The paper by **Hafeez Kabini et al about 'Emotion Based Music Player' [1]** addressed the problem of the existing methods typically handle only deliberately displayed and exaggerated expressions of prototypical emotions despite the fact that deliberate behavior differs in visual appearance, audio profile, and timing from spontaneously occurring behavior, by taking efforts to develop algorithms that can process naturally occurring human affective behavior have recently emerged.

They introduced and surveyed these recent advances and discussed human emotion perception from a psychological perspective. They examined available approaches to solving the problem of machine understanding of human affective behavior, and discuss important issues like the collection and availability of training and test data.

The „mind" is a term that has always attracted scientists towards understanding it in a wholesome manner. The most natural way to express emotions is using facial expressions. We humans, often use

nonverbal cues such as hand gestures, facial expressions, and tone of the voice to express feelings in interpersonal communications.

In the paper by **Nikhil Zaware, Tejas Rajgure, Amey Bhadang, D.D. Sakpal on "Emotion Based Music Player" [2]** stated that it is very time consuming and difficult to create and manage large playlists and to select songs from these playlists. Thus, it would be very helpful if the music player itself selects a song according to the current mood of the user using an application to minimize the efforts of managing playlists.

In their paper, they stated a way to automatically detect the mood of the user and generate playlist of songs which is suitable for the current mood. The image is captured using webcam and that image is passed under different stages to detect the mood or emotion of the user.

The application is thus developed in such a way that it can manage content accessed by user, analyses the image properties and determine the mood of the user.

In another paper **'An accurate algorithm for generating a music playlist based on facial expression' [3**] the proposed algorithm revolves around an automated music recommendation system that generates a subset of the original playlist or a customized playlist in accordance with a user's facial expressions. It is composed of three main modules: Facial expression recognition module, Audio emotion recognition module and a System integration module.

Facial expression recognition and audio emotion recognition modules are two mutually exclusive modules. The system integration module maps the two subspaces by constructing and querying a meta-data file. Implementation and experimentation in this paper was carried out using MATLAB. Testing was first carried out using Cohn- Kanade dataset and then to achieve real time performance, self-annotated dataset was used.

In **'A Media Player which operates depending on Human Emotions' [4],** The proposed system consists of an interface to determine the human gestures including the facial emotions. They also provided a media player that can be controlled by the human gestures. They referred to textured face. Face was captured by webcam and it was detected by Viola-Jones algorithm. Then, using an edge detection the unwanted texture details were eliminated. The algorithm used for edge detection was Canny Edge Detection.

As the algorithm uses double thresholding, edges stronger than a certain value (threshold) were selected. After this, they performed edge tracking by Hysteresis. The image size used by the system was 60x60, thus image resizing was required.

In this paper, **'Facial Expression Analysis' [5],** they show the automation of the entire process of facial expression and give the solution with high accuracy that would be enormously beneficial for the system as diverse as medicine and communication.

This paper shows a lot of progress in past few years and faced a lot of problems to resolved the problem but then also they have reached to that point to showcase our skills in this paper they have used the machine learning technology in this research paper.

In **'Emotion Based Mood Enhancing Music Recommendation' [6],** they proposed an android application which gives user a list of song based on user's current emotion. At the start of the application if the user has internet connection it will provide the song from online or otherwise it played song from the device memory.

If user want emotion-based music it will immediately start camera activity and capture image. Next it will upload the image to server and detect face using Viola Jones algorithm and will give to FisherFace algorithm for recognition of emotion.

Detected emotions are send to the user's device and are also send to the server for fetching the list of song. At last, it will generate the playlist based on emotion recognized. They used 450 images to train the classifier and tested its accuracy. The images used are from CK+ database.

In **'Detection and Recognition of Human Emotion using Neural Network' [7],** they developed a robust system which can detect and recognize the human emotion from live feed consisting mainly of two stages- facial detection which is done by extraction of Haar-Cascade features of a face using viola Jones

algorithm and the second stage is emotion recognition which is done by Deep Neural Network. They used FER2013 dataset having thousands of images with all types of emotion for their system. For training their system they used the 9000 images from the FER2013 dataset.

In **'Moody Player: A Mood based Music Player' [8],** they have proposed a bimodal emotion recognition system with the combination of facial expressions and speech signals.

The models obtained from a bimodal corpus with six acted emotions and ten subjects were trained and tested with different classifiers, such as Support Vector Machine, Naive Bayes and K-Nearest Neighbor. Its result reveals that facial expression gives better performance compared to speech and combination of both also gives better performance with SVM classifier.

In **'Histograms of Oriented Gradients for Human Detection' [9],** in their paper reviewed existing and gradient based descriptors, they showed experimentally that grids of Histograms of Oriented Gradient descriptors significantly outperform existing feature sets for human detection they concluded that fine-scale gradients fine orientation binning, relatively coarse spatial binning, an high quality local contrast normalization in overlapping descriptor blocks are all important for good results.

In **'Facial Expression Recognition based on Local Region Specific Features and SVM' [10],** the input images are selected from the training set. After this Landmark detection & Local representation will be done. By using LBP (Local Binary Pattern) algorithm, Local regions LBP features and Local Regions NCM (Normalized Centre Movement) features are extracted.

Both the extracted features are added together and all these are passed to the SVM (Support Vector Machine) classifier.

CK+ dataset consisting of 593 sequences of different emotions from 123 subjects. Only 327 out-off 593 sequences were given label for the human facial expression. They used at least two peak expression frames for anger, fear, sadness and one peak expression frame for disgust, happy, surprise. The system was trained with 6 types of facial expression.

IV. PROPOSED TECHNIQUE

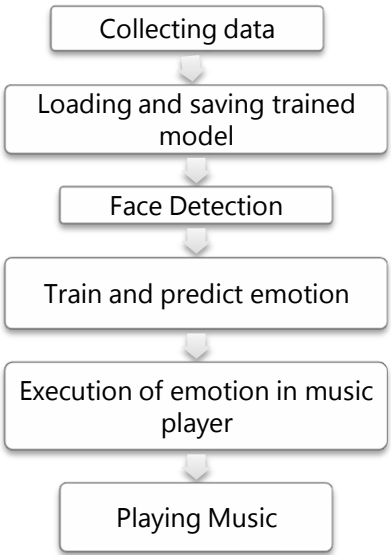

Fig 1. Proposed Technique.

V. IMPLEMENTATION

The technology used in Emotion based music player is Python which is used in the execution of code. The libraries that are used in the project are Open CV Haarcascade and Eeel.

Open CV is a library of Python bindings to solve computer vision problems and it also used in image Normalization, Edge Detection and several exceptional features. Haar Cascade is basically a classifier which is used to detect the object the object for which it has been trained for, from the source. Eeel library is basically used to connect python to browser. Apart from python and the libraries of Python other technologies that are used in the project are HTML, CSS, JavaScript and jQuery.

1. Collecting Data:

Facial expression detection in Fisher face works with the help of trained models. Reason behind this is to allow user to take dataset according to their use. Suppose if we take a huge amount of dataset of around 25-30k it will give nice accuracy no doubt but if the situation is like that the user of the devices, are a few people.

Now in such condition if we take some precise dataset with around 400-450 images as input related to the user then it will also give good accuracy with

the benefit of less amount of dataset and less storage on memory to operate.
As well as small memory of data give output fast which result in quick response time. Here we first tried with Cohn-Kanade dataset then we made some classification in the as our need make it to train our model.

2. Loading and Saving Trained Model:
For training, we have used Fisher face method of cv2 library.

```
fishface = cv2.face.FisherFaceRecognizer_create()
```

Fig 2. To train purpose.

For training data model, we have made a python code which grab all the classified images from folders and map it with its emotion. These data we at an instance stored in dictionary and then use train method to train model.

```
fishface.train(training_data, np.asarray(training_label))
```

Fig 3. To train data model.

To save the model for later use we have implemented. save method.

```
fishface.save("model1.xml")
```

Fig 4. To save model.

Now at the detection time first we have load model in memory using. read method.

```
fishface.read("model.xml")
```

Fig 5. To read model.

Prediction of result is based on the prediction and confidence value which. predict method return.

```
pred, conf=fishface.predict(facedict[i])
```

Fig 6. To predict result.

3. Haarcascade Model:
Haarcascade model is precise face detection trained model which is provided by Open-cv. It returns the co-ordinates in terms of (x, y) at (left, bottom) of face frame and its width and height from those co-ordinates.

```
clahe_image=grab_face()
face=facecascade.detectMultiScale(clahe_image, scaleFactor=1.1, minNeighbors=15,
minSize=(10, 10), flags=cv2.CASCADE_SCALE_IMAGE)
```

Fig 7. Haarcascade model.

As here in the. detectMultiScale () method it is capable of detect multiple faces and it return an array of all the faces(co-ordinates) as an element. The arguments have set according to the threshold what we need for our checking purpose. We have set it such like it doesn't affect our model accuracy.

4. Fisher face ML Algorithm:
Fisher face algorithm is an algorithm which work on the basis of LDA and PCA concepts. Linear discriminate analysis (LDA) is a supervised Learning method of machine learning. Now supervised Learning is that where we use such data whose answer is also given to the model to learn it.

It works on the concept of dimensionality reduction. Which reduce the execution time among classification. Principal Component Analysis (PCA) is a one kind of conversion from correlated variables to uncorrelated in the form of mathematical values. It is mostly used for the observing data and from that by some probabilistic calculation generate models.

The flow of Fisherface is like it takes classified images then it will reduce the dimension of the data and by calculating its statistical value according the given categories it stores numeric values in .xml file. While prediction it also calculates the same for given image and compare the value with the computed dataset values and give according result with confidence value.

5. Resizing Images:
Whatever the image we have chosen for dataset it mostly related to the size which can give a precise output. The size is chosen such like the model can able to easily distinguish face from image by haarcascade model. And the size what we get from real time scan is not always same as data (very less difference) so, we resize it to the exact model data size. In our case we have chosen 350*350. Here In this method, we have implemented the cropping of image by given parameters of haarcascade by clahe_image [] and use of cv2's method. resize () to the given size. Finally, we have stored those images in dictionary and after some count (=10) take it to check result.

6. Gray Scaling Images:

It was the need for the method and because of its contrast and shaded face, it results in benefit for algorithm to get output.

7. Face Detection:

Fig 8. To detect face.

As the given in the code grab_face () methods use to get the images and do all operation and finally return cropped, grayed face value in dictionary.

VI. RESULTS & ANALYSIS

Testing and implementation is performed using Open CV version 3.0 on Windows10, 64bit operating system and Intel i7 core processor. Facial expression extraction is done on both user independent and dependent dataset.

A dataset consisting of facial image of 10 individuals was selected for user independent experiment and dataset of 4 individuals was selected for user dependent experimentation. We have taken four expressions into record, which includes happy, sad, anger, anxiety.

When the camera of the system takes our expression into record it detects the emotion and Plays music according to the track designed to the particular emotion in our project. This helps the human to reduce their efforts on searching for the song and playing it. In this way it makes the human effort less and plays music according to their mood and helps the music lovers.

Happy, Neutral, Sad and Angry. These are the 4 categories which we have choose for this Emotion based music player project. Based on the mood of the person different songs will be played in random order syncing with the mood of the person.

Table 1. Estimation of time of modules.

Sr	Module	Time Taken(sec)
1.	Face Detection	2.9
2.	Landmark points extraction	3.8
3.	Classification	1.5
4.	Emotions	1.1

Table 2. Estimated time for music player module.

Sr	Module	Time taken(sec)
1.	Emotion extraction	9.5
2.	Total time for running	1.1

Table 3. Accuracy of Emotions.

Emotions	Accuracy
Happy	95%
Sad	93%
Neutral	89%
Angry	94%

VII. CONCLUSION

The framework along these lines goes for giving the Windows working framework clients with a less expensive, extra equipment free and exact feeling-based music framework. The Emotion Based Music Framework will be of extraordinary favorable position to clients looking for music dependent on their mind-set and enthusiastic conduct. It will help diminish the looking time for music in this manner diminishing the superfluous computational time and, in this manner, expanding the generally speaking precision and effectiveness of the framework.

The framework won't just diminish physical pressure yet will additionally go about as a help for the music treatment frameworks what's more, may likewise help the music specialist to therapize a patient. Likewise, with its extra highlights.

The future extension in the framework would to plan a component that would be useful in music treatment and give the music specialist the assistance expected to treat the patients experiencing disarranges like mental pressure, uneasiness, intense wretchedness and injury. The proposed framework likewise will in general evade in future the erratic outcomes delivered in extraordinary terrible light conditions and exceptionally poor camera goals.

VIII. ACKNOWLEDGEMENTS

I would like to thank Dr. Rajkumar R for helping me to find out this interesting topic.

REFERENCES

[1] Hafeez Kabini, Sharik Khan, Omar Khan, Shabana Tadvi "Emotion Based Music Player "International Journal of Engineering Research and General Science, Volume 3, Issue 1, 2015.

[2] Nikhil Zaware, Tejas Rajgure, Amey Bhadang, D.D. Sakpal "Emotion Based Music Player" International Journal of Innovative Research & Development, Volume 3, Issue 3, 2014.

[3] AnukritiDureha "An accurate algorithm for generating a music playlist based on facial expressions "International Journal of Computer Applications, Volume 100-No.9, 2014.

[4] Harshala Chaudhari, Amrapali Waghmare, Reshma Ganjewar, Dr. Abhijit Banubakode, International Journal of Advanced Research in Computer and Communication Engineering Vol. 4, Issue 10, May 2015, A Media Player which operates depending on HumanEmotions.

[5] Tian, Y.L., Kanade, T. Cohn, J.F. (2005). Facial Expression Analysis, In Handbook of Face Recognition, Li, S.Z. Jain, A.K., (Eds.), pp. 247-276, Springer, New York, USA.

[6] Aurobind V. Iyer, and Viral Prasad," Emotion Based Mood Enhancing Music Recommendation" in 2017 2nd IEEE International Conference on Recent Trends in Electronics Information & Communication Technology (RTEICT), India, pp. 1573-1577, May 19-20,2017.

[7] J. Jayapradha Soumya Sharma and Yash Dugar," Detection and Recognition of Human Emotion using Neural Network", in International Journal of Applied Engineering Research, Vol. 13, No. 8 (2018), pp. 6472-6477.

[8] A. R. Patel, A. Vollal, P. B. Kadam, S. Yadav, R. M. Samant, "MoodyPlayer: A Mood basedMusic Player".

[9] Histograms of Oriented Gradients for Human Detection Navnee Dalal and Bill TriggsINRIA Rhone-Alps,655 Avenue de l" Europe, Montbonnot 38334, France.

[10] Deepak Ghimire, Sung Wan Jeong, Joonwhoan Lee, Sang Hyun Park," Facial Expression Recognition based on Local Region-Specific Features and SVM" in Multimedia Tools and Applications, Vol. 76, Issue 6, pp. 7803–7821, March 2017.