

Music Recommendation System Leveraging Facial Emotion

¹Gayatri Kale, ²Smita Birajdar, ³Divya Suryawanshi, ⁴Dr. Sarika Khope

^{1,2,3}Student, Electronics and Telecommunication Engineering, G H Raisoni College of Engineering and Management, Pune, Maharashtra, India

⁴Professor, Electronics and Telecommunication Engineering, G H Raisoni College of Engineering and Management, Pune, Maharashtra, India

Abstract - Music plays a significant role in improving and elevating one's mood as it is one of the important source of entertainment and inspiration to move forward. Recent studies have shown that humans respond as well as react to music in a very positive manner and that music has a high impact on human's brain activity. Now a day, people often prefer to listen to music based on their moods and interests. This work focuses on a system that suggests songs to the users, based on their state of mind. In this system, computer vision components are used to determine the user's emotion through facial expressions. Once the emotion is recognized, the system suggests a song for that emotion, saving a lot of time for a user over selecting and playing songs manually. Conventional method of playing music depending upon the mood of a person requires human interaction. Migrating to the computer vision technology will enable automation of such system. To achieve this goal, an algorithm is used to classify the human expressions and play a music track as according to the present emotion detected. It reduces the effort and time required in manually searching a song from the list based on the present state of mind of a person. The expressions of a person are detected by extracting the facial features using the PCA algorithm and Euclidean Distance classifier. An inbuilt camera is used to capture the facial expressions of a person which reduces the designing cost of the system as compared to other methods. The results show that the proposed system achieves up to 84.82.

Keywords: Recommender System, PCA Algorithm, image processing, Artificial intelligence, Classification.

I. INTRODUCTION

Facial expressions are one of the natural means to communicate the emotions and these emotions can be used in entertainment and Human Machine Interface (HMI) fields. In today's world, with the advancements in the areas of technology various music players are deployed with features like reversing the media, fast forwarding it, streaming

playback with multicast streams. Although these features satisfy the basic requirements of the user, yet one has to manually surf for the song from a large set of songs, according to the current circumstance and mood [3] [4]. This is a time consuming task that needs some effort and patience. The main objective of this work is to develop an intelligent system that can easily recognize the facial expression and accordingly play a music track based on that particular expression/emotion recognized. The seven universally classified emotions are Happy, Sad, Anger, Disgust, Fear, Surprise and Neutral. The algorithm that is used in developing the present system is Principal Component Analysis (PCA) which utilizes Eigen faces to extract the facial features. The designed algorithm is very efficient due to less computational time taken hereby increasing the performance of the system. This work finds its applications in various domains like Human Computer Interaction (HCI), therapeutic approach in health care etc. Most of the time the digital music is sorted and put to gather based on attributes such as artist, genre, albums, language, popularity and so on. Many of the available online music streaming services recommend music based on user's preferences and his previous music listening history that employ content based and collaborative filtering recommendations. But these recommendations may not suite the current mood of the user. The manual classification of songs by learning user's preference of emotion is a time consuming task. So, recommendations can also be achieved using the physiological and emotional status of the user which are mainly captured from the user's facial expression, gestures, pulse rate, movement, speech/text interactions etc. Several work is carried to detect emotions using facial landmarks to extract the features. Nguyen et al. [11] detected three kinds of emotions namely positive, negative and blank using 68 facial landmarks with an accuracy of 70.65.

II. LITERATURE SURVEY

1. Paper Name: Research on Automatic Music Recommendation Algorithm Based on Facial Micro Expression Recognition.

Author: Ziyang Yu¹, Mengda Zhao¹, Yilin Wu¹, Peizhuo Liu¹, Hexu Chen¹.

Abstract: In recent years, with the development and application of big data, deep learning has received more and more attention. As a deep learning neural network, convolutional neural network plays an extremely important role in face image recognition. In this paper, a combination of micro-expression recognition technology of convolutional neural network and automatic music recommendation algorithm is developed to identify a model that recognizes facial micro-expressions and recommends music according to corresponding mood. The facial micro-expression recognition model established in this paper uses FER2013 with a recognition rate of 62.1.

2. Paper Name: EMOSIC- An Emotion Based Music Player for Android.

Author: Karthik Subramanian Nathan, Manasi Arun and Megala S Kannan.

Description: Music plays a very important role in human's daily life and in the modern advanced technologies. Usually, the user has to face the task of manually browsing through the playlist of songs to select. Here we are proposing an efficient and accurate model, that would generate a playlist based on current emotional state and behavior of the user. Existing methods for automating the playlist generation process are computationally slow, less accurate and sometimes even require use of additional hardware like EEG or sensors. Speech is the most ancient and natural way of expressing feelings, emotions and mood and its processing requires high computational, time, and cost. This proposed system based on real-time extraction of facial expressions as well as extracting audio features from songs to classify into a specific emotion that will generate a playlist automatically such that the computation cost is relatively low.

3. Paper Name: Emotion aware Smart Music Recommender System using Two Level CNN.

Author: Krupa K S, Ambara G.

Description: Music plays a significant role in improving and elevating one's mood as it is one of the important source of entertainment and inspiration to move forward. Recent studies have shown that humans respond as well as react to music in a very positive manner and that music has a high impact on human's brain activity. Nowadays, people often prefer to listen to music based on their moods and interests. This work focuses on a system that suggests songs to the users, based on their state of mind. In this system, computer vision components are used to determine the user's emotion through facial expressions and chatbot interactions. Once the emotion is recognized, the system suggests a song for that emotion, saving a lot of time for a user over selecting and playing songs manually.

4. Paper Name: Music Recommendation Based on Color.

Author: Kajornsak Kittimathaveenan, Salisa Mahatanarat.

Abstract: This study presents an alternative way in choosing songs based on a selection of colors through Color-to-Music application. There were three stages of this study: The first stage was the preparation music library of the association between color and emotion; and the association between music and emotion. Library data used for the Hue, Saturation, and Value (HSV) color model creation were: Hue to represent musical instruments, Saturation referred to tempo, and Value was key (pitch). Second stage was to create two types of graphical user interface (GUI) for color selection. The last stage was to collect data from 120 participants participating in the trials. This study focused on two questions. First was the accuracy rate of a recommended song that matched selected colors. Second was to find the most suitable GUI that provides the highest accuracy rate to the recommendation of songs. The tests were divided into two groups: test A and test B. As for test A, participants started a trial by choosing color from the application; while test B would start by selecting an initial emotion, and then choosing colors to match the chosen emotions. The results showed that the overall accuracy rate of test A is higher than test B and the slider GUI has the highest accuracy rate.

III. SYSTEM DESIGN

The system architecture for the proposed system is given in fig. The input image is loaded into the system in .jpg format. Then each image undergoes preprocessing i.e. removal of unwanted information like background color, illumination and re sizing of the images. Then the required features are extracted from the image and stored as useful information. These features are later added to the classifier where the expression is recognized with the help of Euclidean distance. Minimum the value of the distance calculated, the nearest match will be found. Finally, a music track will be played based on the emotion detected of the user.

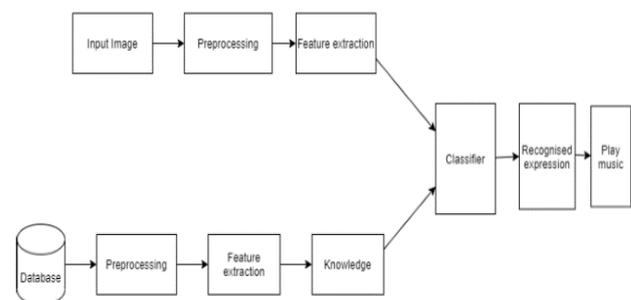


Figure 1: System Architecture

Steps involved to design the system, training dataset and test images are considered for which the following procedures are applied to get the desired results. The training set is the

raw data which has large amount of data stored in it and the test set is the input given for recognition purpose. The whole system is designed in 5 steps:

1. Image Acquisition

In any of the image processing techniques, the first task is to acquire the image from the source. These images can be acquired either through camera or through standard datasets that are available online. The images should be in .jpg format. The images considered here are user dependent i.e. dynamic images. The number of sample training images considered here.

2. Pre-processing

Pre-processing is mainly done to eliminate the unwanted information from the image acquired and fix some values for it, so that the value remains same throughout. In the pre-processing phase, the images are converted from RGB to Gray-scale and are resized to 256*256 pixels. The images considered are in .jpg format, any other formats will not be considered for further processing. During pre-processing, eyes, nose and mouth are considered to be the region of interest. It is detected by the cascade object detector which utilizes Jones-Viola algorithm.

3. Facial Feature Extraction

After pre-processing, the next step is feature extraction. The extracted facial features are stored as the useful information in the form of vectors during training phase and testing phase. The following facial features can be considered “Mouth, forehead, eyes, complexion of skin, cheek and chin dimple, eyebrows, nose and wrinkles on the face”. In this work, eyes, nose, mouth and forehead are considered for feature extraction purpose for the reason that these depict the most appealing expressions. With the wrinkles on the forehead or the mouth being opened one can easily recognize that the person is either surprised or is fearful. But with a person’s complexion it can never be depicted. To extract the facial features PCA technique is used.

4. Expression Recognition

To recognize and classify the expressions of a person Euclidean distance classifier is used. It gets the nearest match for the test data from the training data set and hence gives a better match for the current expression detected. Euclidean distance is basically the distance between two points and is given by “(3.1)”. It is calculated from the mean of the eigenfaces of the training dataset. The training images that correspond to various distances from the mean image are labeled with expressions like happy, sad, fear, surprise, anger,

disgust and neutral. When the Euclidean distance between the eigenfaces of the test image and mean image matches the distance of the mean image and eigenfaces of the training dataset the expression is classified and named as per the labeled trained images. Smaller the distance value obtained, the closest match will be found. If the distance value is large enough for an image, then the system has to be trained for that individual. The equation measure using Euclidean distance.

5. Play Music

The last and the most important part of this system is the playing of music based on the current emotion detected of an individual. Once the facial expression of the user is classified, the user’s corresponding emotional state is recognized. A number of songs from various domains pertaining to a number of emotions is collected and put up in the list. Each emotion category has a number of songs listed in it. When the user’s expression is classified with the help of PCA algorithm, songs belonging to that category are then played.

IV. RESULTS AND DISCUSSIONS

The system detects the mood in real time and a playlist is displayed for that emotion accurately. It is able to fetch and play the songs from the recommended playlist in the presence of a stable internet connection. The playlist displays the songs to the user in list view, and when the user selects a song to be played, the song is played in a media player, that comes with all the necessary buttons such as play, pause, shuffle, next and previous, along with a seek bar. For “angry”, “Fear”, “disgust” and “surprise” moods, devotional, motivational and patriotic songs are suggested to the user. Hence, the user is also provided with mood improvement.

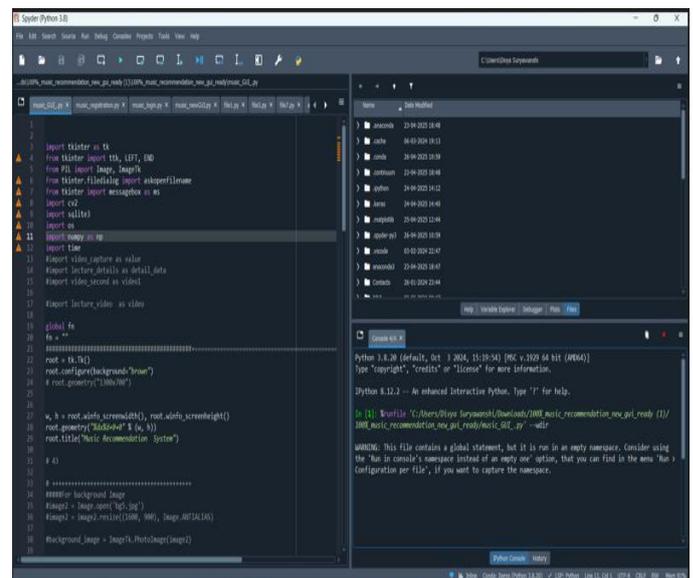


Figure 2: Code Snapshot

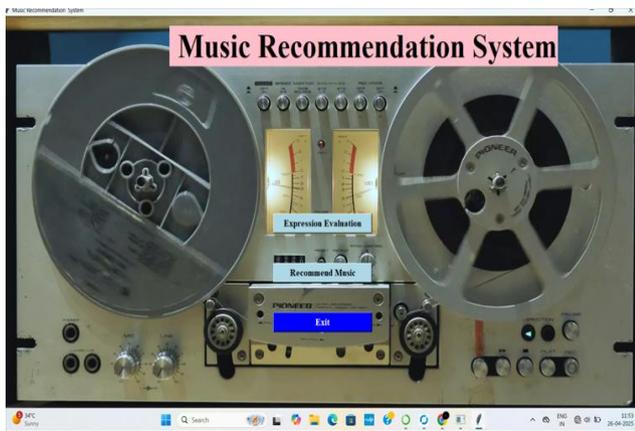


Figure 3: Result

V. CONCLUSION

The proposed work presents facial expression recognition system to play a song according to the expression detected. It uses PCA approach to extract features, and Euclidean distance classifier classifies these expressions. In this work, real images i.e. user dependent images are captured utilizing the in-built camera. The final result shows the accuracy level obtained is upto 84.82.

Music recommender system is a system which learns from the users past listening history and recommends them songs which they would probably like to hear in future.

REFERENCES

- [1] S L Happy and Aurobinda Routray, "Automatic Facial Expression Recognition using Features of salient Facial Patches," in IEEE Trans. On Affective Computing, January-March 2015, pp. 1-12.
- [2] Hafeez Kabani, Sharik Khan, Omar Khan and Shabana Tadvi, "Emotion based Music Player," Int. J. of Eng. Research and General Sci., Vol. 3, Issue 1, pp. 750-756, January-February 2015.
- [3] Li Siquan, Zhang Xuanxiong. Research on Facial Expression Recognition Based on Convolutional Neural Networks [J]. Journal of Software, 2018, v.17; No.183 (01): 32-35.
- [4] Hou Yuqingyang, Quan Jicheng, Wang Hongwei. Overview of the development of deep learning [J]. Ship Electronic Engineering, 2017, 4: 5-9.
- [5] Liu Sijia, Chen Zhikun, Wang Fubin, et al., Multi-angle face recognition based on convolutional neural network [J]. Journal of North China University of Technology (Natural Science Edition), 2019, 41 (4): 103-108.
- [6] Li Huihui. Research on facial expression recognition based on cognitive machine learning [D]. Guangzhou: South China University of Technology, 2019.

- [7] Li Yong, Lin Xiaozhu, Jiang Mengying., Facial expression recognition based on cross-connection LeNet-5 network [J]. Journal of Automation, 2018,44 (1): 176-182.
- [8] Yao L S, Xu G M, Zhap F., Facial Expression Recognition Based on CNN Local Feature Fusion [J]., Laser and Optoelectronics Progress, 2020, 57(03): 032501.
- [9] Xie S, Hu H. Facial expression recognition with FRR-CNN [J]. Electronics Letters, 2017, 53 (4): 235-237.
- [10] Zou Jiancheng, Deng Hao., An automatic facial expression recognition method based on convolutional neural network [J]. Journal of North China University of Technology, 2019, 31 (5): 51-56.

AUTHORS BIOGRAPHY



Gayatri Kale, Student, Electronics and Telecommunication Engineering, G H Raisoni College of Engineering and Management, Pune, Maharashtra, India.



Smita Birajdar, Student, Electronics and Telecommunication Engineering, G H Raisoni College of Engineering and Management, Pune, Maharashtra, India.



Divya Suryawanshi, Student, Electronics and Telecommunication Engineering, G H Raisoni College of Engineering and Management, Pune, Maharashtra, India.



Dr. Sarika Khope, Professor, Electronics and Telecommunication Engineering, G H Raisoni College of Engineering and Management, Pune, Maharashtra, India.



Citation of this Article:

Gayatri Kale, Smita Birajdar, Divya Suryawanshi, & Dr. Sarika Khope. (2025). Music Recommendation System Leveraging Facial Emotion. *International Research Journal of Innovations in Engineering and Technology - IRJIET*, 9(5), 447-451. Article DOI <https://doi.org/10.47001/IRJIET/2025.905050>
