"Facial Emotion based Music Recommendation using Convolutional Neural Networks"

Krishna Bhalekar, Karan Nilesh Dound, Abhishek Balaji Panchgalle

Computer Engineering , Savitribai Phule Pune University ,Pune, Maharashtra ,India.

ABSTRACT

This work presents a music recommendation system that suggests songs based on the user's emotional state. The system uses computer vision techniques to detect the user's emotions through facial expressions, and a suitable song is recommended accordingly. The proposed system automates the traditional manual process of selecting music based on mood, reducing time and effort. The algorithm uses Haar Cascade and CNN algorithms to detect facial expressions and select an appropriate music track. The inbuilt camera reduces system design costs. The system is based on recommender systems, convolutional neural networks, deep learning, image processing, artificial intelligence, and classification**.**

Keywords: Emotion recognition, CNN, Facial expression, Semantic analysis, Machine Learning

I. INTRODUCTION

Facial expressions can be used as a natural means to communicate emotions and have potential applications in the entertainment and human-machine interface domains. While current music players offer features such as reversing, fast forwarding, and streaming playback, users still need to manually search for a song based on their current mood and circumstance. To address this, an intelligent system is proposed that can recognize facial expressions and play a music track accordingly. The system uses the Haar Cascade algorithm and Eigen faces to efficiently extract facial features for emotion recognition. This approach can be applied to various domains, including human-computer interaction and healthcare. While current music streaming services recommend music based on user preferences and listening history, this approach uses physiological and emotional cues captured from facial expressions, gestures, pulse rate, movement, and speech/text interactions to recommend music based on the user's current emotional status. By employing a CNN-based approach to analyse facial expressions, the proposed system can provide efficient and accurate music recommendations.

II.METHODS AND MATERIAL

Steps involved to design the system 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. Images taken as input are user based 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.0020 Features are detected by the movement "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 recognise 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 are annotated with affective labels corresponding to different degrees of detachment from the mean image. These labels include affective states such as happiness, sadness, dread, thunderbolt, 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 labelled 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. After classifying the facial expression of the user, the corresponding affective state of the user is recognized.. That sounds like a great way to organize a music collection! Having songs categorized by emotion can help you select music that matches your mood or helps you shift your mood to a desired emotional state. When the user's expression is classified with the help of CNN algorithm, songs belonging to that category are then played.

III. RESULT AND DISCUSSION

A. Mathematical Model

Let S be the whole system S= {I ,P,O}

I-input

P-procedure

O-output

Input (I)

I= { live camera}

Where,

Emotions -> happy sad neutral

Procedure (P),

P= {I, Using I System perform type of emotions.}

Output(O)-

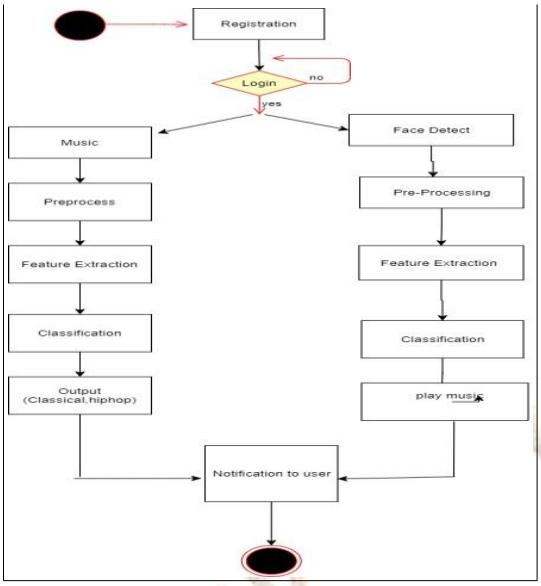
O= {System detect type of emotions, happy sad or neutral}

Convolutional neural network algorithm

A Convolutional Neural Network (CNN) is a class of deep learning algorithms that is specifically designed to handle complex visual processing tasks such as image and video recognition. At a high level, CNNs consist of several layers of interconnected neurons, each of which performs a specific function in the network's overall processing pipeline.The role of the CNN algorithm would be to extract facial features from the input image, specifically the emotional expressions on the human face, and use these features to make predictions about the emotion being displayed.

To achieve this, the CNN would be trained on a dataset of images with labeled emotions, such as happy, sad, angry, and so on. The training process involves iteratively adjusting the network's parameters to minimize the difference between its predicted emotions and the true emotions in the training data.Once the CNN has been trained, it can be used to predict the emotions in new input images. The input image in your case would be a live video stream from a camera capturing a person's face. The CNN would process each frame of the video stream and output the predicted emotion for that frame.Based on the predicted emotion, your system would recommend a suitable song to the user. The specific recommendation algorithm and song dataset you use would depend on your project's design.

B. FIGURES AND TABLES



System Architecture

IV. CONCLUSION

The proposed work presents facial expression recognition system to play a song according to the expression detected and also classify music Type. It uses CNN 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.

V. REFERANCE

1. Music Player," Int. J. of Eng. Research and General Sci., Vol. 3, Issue 1, pp. 750-756, January- February 2015.
2. Li Siquan, Zhang Xuanxiong. Research on Facial Expression Recognition
3. 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.