

Audio Recognition System Based on Emotion Using Deep Learning and AI

Swetha Balachandran^{1*}, J. Sharika Mohan², M. Jenefer³, M. Shanmugha Priya⁴

^{1,2,3,4}Department of Computer Science and Engineering, CMS College of Engineering and Technology, Coimbatore, India

Abstract: Human emotions are designed to help people understand each other and share their feelings and intentions. Emotions are expressed through both verbal and facial expressions. Music has a significant impact on an individual's life. This is an innovative method that assists the user in automatically generating a playlist of music based on the user's mood. A web camera is used in this system to identify the user's mood by looking at their facial expressions. Once the mood is identified, the algorithm offers a playlist for that mood, saving the user a lot of time over manually selecting and enjoying songs. It also includes a basic plan for collecting user information such as playlist creation, emotion classification, and interest level, as well as reorganizing the playlist.

Keywords: Audios, Camera, Convolution Neural Network, Emotion, playlist Generation.

1. Introduction

The face is an important component of a person's body, and it is especially important in determining a person's mood. Directly employing a camera to extract the needed input from the face is currently possible. This input will be used to extract data in order to determine a person's mood. This data will be used to generate a list of songs that correspond to the "mood" determined from the previous input. This eliminates the time consuming job of manually categorizing or organizing songs into distinct playlists based on a person's emotional preferences. Convolution Neural Networks are used to perform image-related tasks such as image classification and object detection (CNNs). Furthermore, CNNs are commonly utilized as black boxes, with their learnt features hidden, making it difficult to strike a compromise between classification accuracy and extraneous factors. To validate the features learned by the CNN, real-time visualization of the guided-gradient back-propagation is used. The goal of an emotion-based audio player that uses Deep Learning and AI is to scan and analyze information before creating a playlist based on the parameters provided. It involves audio feature extraction and classification in order to generate a list of songs in the same genre or comparable sounding songs. The application is thus designed in such a way that it will manage the content that the user accesses, evaluate image qualities, and verify the user's mood, supported by mp3 file properties so that they can be supplemented into appropriate

playlists in step with the user's emotions.

2. Literature Survey

1. Face expressions are categorized into two broad types by Zheng et al. to achieve feature recognition: Appearance-based feature extraction and Geometric based feature extraction.
2. Renuka R. Londhe suggested a statistically sound and efficient method for assessing retrieved facial traits. The analysis of variations in curvatures on the face and intensities of corresponding pixels in images was the main focus of the work. The classification of retrieved features into 6 major universal emotions, including anger, disgust, fear, joy, sorrow, and surprise, was done using Artificial Neural Networks (ANN). The researchers employed a Scaled Conjugate Gradient back-propagation algorithmic approach in conjunction with a two-layered feed-forward neural network to achieve a 92.2 percent recognition rate.
3. Thayer offered a useful 2-dimensional (Stress vs. Energy) model displayed on two axes with emotions represented by a two-dimensional coordinate system, laying on either the two axes or the four quadrants produced by the two-dimensional plot. Jung Hyun Kim's work examined and evaluated the music mood tags and A-V values of a total of 20 individuals, and based on the findings, the A-V plane was divided into eight regions (clusters) that depict mood using a data mining effective k-means clustering technique.

3. Existing System

The user must manually categories the songs in accordance with various emotions using the tools available inside the existing audio players included in computer systems. In previous audio players, the user had to actively scroll through his playlist and select music that would help him relax and cope with his emotions. Various audio players are developed in today's world, with ever-increasing advancements in the field of multimedia system and technology, with options such as quick forward, reverse, variable playback speed (seek & time compression), local playback, streaming playback with multicast streams, volume modulation, genre classification, and so on. Despite the fact that these alternatives meet the user's

*Corresponding author: swethabalachandran11@gmail.com

needs, the user is still faced with the burden of manually scrolling through the playlist of songs and selecting songs based on his current mood and behavior.

4. Proposed System

Here, we'd like to suggest an emotion-based music player that plays songs dependent on the user's mood. Its goal is to provide the user with the most popular music while also allowing them to be aware of their emotions. The player is built around the idea of automating much of the interaction between the audio player and the user. The machine learning technology is used to recognize the emotions. A support vector machine is a supervised learning model with related learning algorithms that examine data for classification and multivariate analysis in machine learning. It establishes a link between the possible outputs' optimum boundaries. We usually use the Olivetti faces coaching dataset, which has 400 faces and their intended values or characteristics. The image of the user is captured by the digital camera. The user's countenance is then extracted from the collected image. The training process entails initializing some random values for our model, such as smiling and not smiling, predicting the output using those values, and then comparing the output to the model's prediction values to ensure that they match the predictions made before.

1) Features

- Users do not have to manually select songs.
- There's no need for a playlist.
- Users are not required to categorize music based on their emotions.

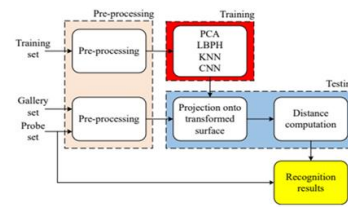


Fig. 1. System architecture

5. Conclusion

The suggested system includes an audio player that can offer a personalized playlist, making it simple for the user to obtain the playlist. The proposed system will make it easier to cut down on music seeking time and extra calculating time, improving the system's overall accuracy and potency. The technology will not only aid patients suffering from anxiety, trauma, acute depression, and mental stress, but it will also aid music therapists treat patients suffering from anxiety, trauma, acute depression, and mental stress.

References

- [1] Mir Kamal Mirnia, A. Habibat, Ninavin suggested an algorithm for classifying face emotions based on eye and lip features has been developed using particle swarm optimization. A new algorithm based on eye and lip features has been developed using particle swarm optimization to categorise face emotions.
- [2] S. Dornbush, K. Fisher, K. McKay, A. Prikhodko and Z. Segall Xpod- A Human Activity and Emotion Aware Mobile Music Player, UMBC Ebiquty, November 2005.
- [3] OpenCV Reference Manual.
- [4] <https://pythonprogramming.net/haar-cascade-face-eye-detection-python-opencv-tutorial/>