JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

EMOTIONRHYTHMS: A REAL-TIME MUSIC PLAYER BASED ON EMOTION RECOGNITION

Yashashri Kamalakant Gosavi

PG Student, Dept. of Information Technology, Patkar-Varde College,
Mumbai, Maharashtra, India

***Abstract*—** EmotionRhythms is an innovative music player that aims to automatically play songs based on the user's current emotions. In today's world, music plays a crucial role in the lives of individuals. Existing music player applications require users to manually search through playlists to find a song that resonates with their current emotional state. People often associate the music they listen to with their emotions, and the large size of song playlists can make it challenging to navigate. A music player that can intelligently organize songs based on the user's emotional state would be highly advantageous. This proposed system would enable users to effortlessly listen to songs that match their current emotions. The proposed system is set to function using four emotions: Happy, Sad, Angry, and Neutral. These emotions will be detected through the use of a machine learning method called CNN (Convolutional Neural Network), which is primarily used for analyzing facial expressions. The system will then play the recommended music from the designated directories.

***Keywords*— Convolutional Neural Networks (CNN), Emotion Recognition, Facial Expression Analysis, Music Recommendation System, Real-Time Emotion Detection, OpenCV**

I. INTRODUCTION

Music is an art form that involves combining different sounds to create a new, pleasant sound. It holds significant importance for most individuals, yet we are selective about the music we listen to. Music is essential for fostering creativity and enhancing children's linguistic abilities. Additionally, music therapy has proven to be beneficial for cancer patients, children with ADHD, and individuals suffering from depression. As a result, many hospitals have incorporated music therapy into their treatment programs. Music is an art form that involves combining different sounds to create a new, pleasant sound. It holds significant importance for most individuals, yet we are selective about the music we listen to. Music is essential for fostering creativity and enhancing children's linguistic abilities. Additionally, music therapy has proven to be beneficial for cancer patients, children with ADHD, and individuals suffering from depression. As a result, many hospitals have incorporated music therapy into their treatment programs. The main objective of EmotionRhythms is to automate the selection of songs based on the users' emotion. To access the functions of EmotionRhythms, users should first login to the web application. When a user accesses the web application, they encounter two options: to either manually select their mood or to enable the webcam for automatic detection. Once the application has determined the user's mood, it automatically initiates playback of a song that matches the identified emotional state. The system employs convolutional neural networks for facial expression recognition, achieving a notable accuracy rate of 95%.

II. LITERATURE REVIEW

1) Within a designated system [1], the testing of functionalities was conducted using Anaconda alongside Python 3.5 software, with the Viola-Jones and Haar cascade algorithms implemented for face detection purposes. Furthermore, the KDEF (Karolinska Directed Emotional Faces) dataset and VGG (Visual Geometry Group) 16 were utilized in tandem with a Convolutional Neural Network (CNN) model, which was developed to attain an accuracy of 88% in face recognition and classification, thus validating the associated performance measures.

2) In another implementation [2], Python 2.7 was utilized in conjunction with the Open-Source Computer Vision Library (OpenCV) and the Cohn-Kanade (CK) and Extended Cohn-Kanade (CK+) datasets, resulting in an accuracy of around 83%.

3) The findings from a cross-database experiment [3] indicated that raw features performed optimally with Logistic Regression in the context of the RaFD (Radboud Faces Database) and Mobile images dataset. The accuracy attained was 66% for the RaFD and

36% for the Mobile images dataset, both utilizing the CK+ dataset for training. In contrast, the addition of extra features, including distance and area, adversely affected the accuracy of the Support Vector Machine (SVM), reducing it from 89%.

4) The field of Music Recommendation Systems has seen significant research efforts. A particular study [4] presents a foundational method for the mood classification of Hindi music, which leverages straightforward features extracted from audio data. The implementation of the MIREX (Music Information Retrieval Evaluation eXchange) mood taxonomy resulted in an average accuracy of 51.56% when evaluated using 10-fold cross-validation.

5) In their research, Renuka R Londhe et al. [5] investigated the variations in facial curvature and the intensity levels of corresponding pixels. The authors utilized Artificial Neural Networks (ANN) to facilitate emotion classification. Furthermore, they proposed multiple strategies for playlist generation.

6) In their study, Zheng et al. [6] delineated two major categories for facial feature extraction: appearance-based and geometric-based feature extraction. The geometric approach specifically involves the extraction of key facial points, such as the mouth, eyes, and eyebrows.

III. EXISTING SYSTEM

The present review highlights several cutting-edge emotion-responsive music player systems, which include, but are not limited to:

- Emo Player: This innovative system utilizes machine learning techniques to interpret facial expressions and physiological metrics, such as heart rate variability, to ascertain the user's emotional state in real-time. It subsequently recommends music tracks that are congruent with the identified emotional states.
- Moodify: By leveraging a user's historical listening patterns and preferences, Moodify applies collaborative filtering and sentiment analysis to assemble playlists that reflect the user's current mood. The system is designed to adapt and improve continuously based on user feedback.
- EmoTunes: This music player focuses on emotions by combining user-generated emotional inputs, such as self-reported mood tags or emoticons, with audio signal analysis to create tailored playlists that align with the user's declared emotional preferences.

IV. PROBLEM STATEMENT

Most of the existing music applications do not take the user’s facial expression as an input parameters for recommending songs. That flaw inspired me to develop an application that focuses on the same. The application I am developing will use AI to detect user facial expressions in real-time and recommend songs based on that. Although I think so focusing on the user's emotions will create my application Uniquely, I am aware that they are not the only contributing factors involved in creating software products. There are plenty corner cases should be taken care of.. One major case is not all users are expressive. Although it is expressive, sometimes due to poor lighting, the application may not take place able to detect the user's face. In such cases, the user can choose mood manually, like how you choose a genre or artist other widely used music applications. The application works for the current four expressions. They are happy, sad, angry and neutral. These are the most common feelings expressed by real life people. I am planning to add more if I have them. The application gives accurate results for existing along with feeling accuracy, providing is also a priority good user experience with attractive user interface.

V. PROBLEM SOLUTION

This research proposes the development of a music application that uniquely plays music according to the user's facial expressions, thereby setting it apart from conventional music applications. Most existing approaches involve either manual music selection, the use of wearable devices, or categorization based on sound characteristics. We propose a shift from these traditional methods towards an automated system. The application is designed as a web-based platform, built using Flask, a web application framework developed in Python. It utilizes MySQL for database management, which stores critical user account information such as email, name, and password. Upon user login, two options are available: users can either manually indicate their mood or permit the webcam to detect it. Following mood detection, the application automatically selects and plays a song that aligns with the identified emotional state. While there are numerous face detection techniques available, we have opted for the Haar cascade algorithm due to its straightforward implementation and real-time processing capabilities. First introduced in 2001, Haar cascades have gained prominence as one of the leading object detection algorithms in the OpenCV library.

VI. METHODOLOGY

A. Mood Detection Module

The following two sections make up this module:

- Face Recognition: The functionality to detect faces in any input image or frame is a crucial capability. The resulting output includes the coordinates of the bounding boxes that encapsulate the identified faces. In this context, the Python library cv2 was employed. This library is the import name for OpenCV-python, which offers "Unofficial pre-built CPU-only OpenCV packages for Python." The conventional method of using OpenCV necessitates a series of intricate steps to compile the module from scratch, which is rendered unnecessary by this method.

- Mood Recognition: Emotions displayed on the face can be categorized into several distinct types: happiness, anger, sadness, neutrality, surprise, fear, and disgust. The Keras module in Python was employed for this classification task. The Haar Cascade algorithm was implemented to detect these emotional expressions. A user's image is captured via a camera or webcam, and this image is then converted to grayscale to optimize the classifier's performance. Once the image is in grayscale, it is processed by the classifier algorithm, which utilizes feature extraction methods to identify the face within the webcam feed. The extracted facial features are then analyzed by a trained neural network to determine the emotion expressed by the user. The images gathered will be utilized to train the classifier, allowing it to recognize and extract facial landmarks from new, unseen images based on previously acquired knowledge. The training of the network is supported by the comprehensive CK dataset, which is instrumental in identifying the emotions expressed by the user.

B. Mood Recommendation Module

A collection of songs, categorized by mood in both Hindi and English, was located on Kaggle. To ensure efficient storage, retrieval, and querying of this data for users, research was undertaken to identify a dependable cloud storage service. Alternatives such as AWS and Google Cloud were evaluated but were ultimately ruled out due to their prohibitive costs and minimal free storage provisions. As a result, this project focuses on recommending songs via YouTube, a widely used online video sharing and social media platform. When a user's emotion is identified, songs that correspond to their facial expressions are recommended, enabling the user to choose their desired track.

C. Integration

The modules responsible for login and signup authentication were implemented using Flask, complemented by a frontend designed with HTML, CSS, and JavaScript. Following user registration, the homepage, which is developed in Python, allows users to either input their emotions directly or opt for a real-time mood detection system. The first component is executed through JavaScript button functionalities, which allow users to click a button that reveals a music playlist specifically curated for the selected mood. This interaction enables users to select a song of their choice for listening. In contrast, the second component is developed using JavaScript in conjunction with the Navigator media Device, a read-only property that returns a Media Devices object, facilitating access to connected media input devices such as cameras and microphones. To detect the user's emotional state, the project employs the cv2 library and the Haar cascade algorithm, utilizing HAAR.js for image feature detection and FILTER.js for video and image processing within a pure JavaScript framework. Furthermore, a Keras model is utilized for emotion detection, linking the recognized emotion to the corresponding music playlist.

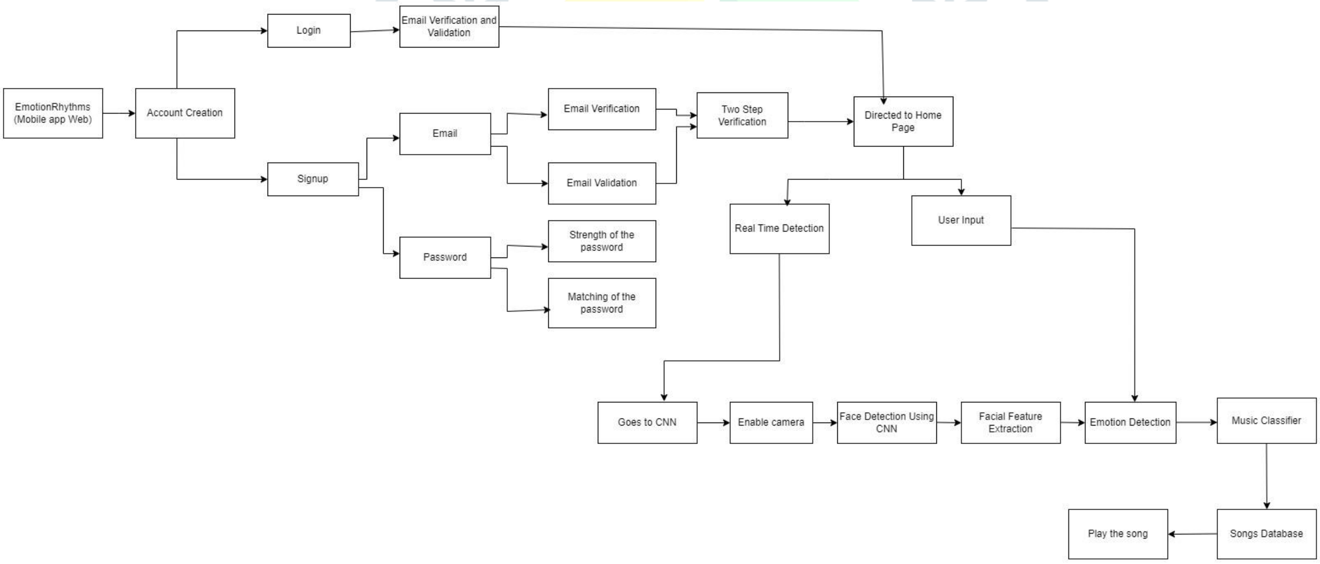

Fig. 1. System Architecture of EmotionRhythms

The system architecture diagram provides a visual representation of the interactions, constraints, and boundaries that exist between components, alongside the general framework of the software system. Upon launching the application, users are presented with a home screen that features three buttons: take a picture, submit input, and play music. When the "take picture" button is selected, the camera interface opens, allowing the user to capture an image. This image is then processed as input for a facial recognition algorithm. If the algorithm does not detect any faces or identifies multiple faces, an appropriate error message is displayed to the user. After the successful identification of a single face, the image is sent as input to the mood detection module.

Upon detection of the user's mood, the system displays the identified emotional state, subsequently activating the "play tunes" button. This functionality allows the user to select and play a song from a curated playlist that aligns with the recognized mood. As depicted in Fig. 1, a set of five emotion buttons becomes visible when the user interacts with the "how do you feel" button.

The user can click any of these buttons to access the relevant playlist. To exit the application, the user simply needs to press the back button.

VII. CONCLUSION

Accurately identifying human emotions or moods poses significant challenges due to the unique facial characteristics of each individual. However, with appropriate facial expressions, it is possible to recognize emotions to some extent. The implementation of a high-resolution camera is essential for the accurate detection of these expressions. The development of this Python-based recognition system is labor-intensive, requiring the integration of multiple technologies. Despite the potential for further refinement, the primary objective of this project—automatically selecting music that aligns with a user's emotional state—has been partially achieved. While it is acknowledged that the system is not flawless, it successfully identifies a range of emotions, including happiness, sadness, anger, neutrality, and surprise. Upon detecting an emotion, the proposed system generates a playlist that corresponds to the identified mood. However, processing large datasets increases both memory and CPU usage, complicating the development process. The goal remains to develop this application as cost-effectively as possible, using a widely accessible platform. Ultimately, this facial emotion-based music recommendation system aims to simplify the tasks of playlist creation and management for users.

VIII. REFERENCES

[1] Dr. Shaik Asif Hussain and Ahlam Salim Abdallah Al Balushi, "A real time face emotion classification and recognition using deep learning model", 2020 Journal. of Phys.: Conf. Ser. 1432 012087.

[2] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar and I.Matthews, "The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion specified expression," 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops, San Francisco, CA, USA, 2010, pp. 94-101, doi:10.1109/CVPRW.2010.5543262.

[3] Raut, Nitisha, "Facial Emotion Recognition Using Machine Learning" (2018). Master's Projects. 632. https://doi.org/10.31979/etd.w5fs-s8wd

[4] Patra, Braja & Das, Dipankar & Bandyopadhyay, Sivaji. (2013). Automatic Music Mood Classification of Hindi Songs.

[5] Londhe RR and Pawar DV 2012 Analysis of facial expression and recognition based On statistical approach International Journal of Soft Computing and Engineering 2.

[6] Zeng Z, Pantic M, Roisman GI and Huang TS 2008 A survey of affect recognition methods Audio, visual, and spontaneous expressions IEEE transactions on pattern analysis and machine intelligence 31 39-58

[7] Ketki R. Kulkarni and Sahebrao B. Bagal, "Facial expression recognition", 2015 International Conference on Information Processing (ICIP).

[8] V. R. Ghule, A. B. Benke, S. S. Jadhav, and S.A. Joshi, "Emotion Based Music Player Using Facial Recognition," vol. 5, no. 2, p. 7, 2007. International Journal of Innovative Research In Computer and Communication, 2017.

[9] Gokul Krishnan K, Parthasarathy M, Sasidhar D and Venitha E, "Emotion detection and music recommendation system using machine learning" in 2018 International Journal of Pure and Applied Mathematics, 2018, vol. 119, pp. 1487-1498.

[10]Shlok Gilda, Husain Zafar, Chintan Soni and Kshitija Waghurdekar, "Smart Music Player Integrating Facial Emotion Recognition and Music Mood Recommendation", presented at the IEEE WiSPNET 2017 conference.