IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 9, Issue 7, July 2021

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.542

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

|| Volume 9, Issue 7, July 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0907135 |

Computer Vision Based Smart Music Player

Nagamani D. R, Abhishek Shrivastava, Madhu B, Channaveer Swamy, Chandan B. S. D, B. K. Paritosh

Department of Computer Science and Engineering, Bangalore Institute of Technology, Bengaluru, India

ABSTRACT: Music is a universal form of entertainment bringing joy and happiness to avid listeners throughout the world. Most music players require the user to select songs manually, leading users to create a playlist of songs for each mood. Though these music players have advanced features like providing lyrics and the ability to tweak listening parameters such as bass, temp among others there is room for improvement. We aim to automate the selection of songs through the system reacting to the user's emotion obtained using his facial expression and classifying them into 4 categories or moods namely - Happy, Angry, Sad, Neutral, saving time that would have been spent entering information manually. To provide a non touch music control experience our music player uses a hand gesture recognition system for all music control operations such as pause, play, skip, increase volume, decrease column and skip a song. Finally if due to the COVID19 pandemic the user is wearing a face mask then we detect it and instruct the user to specify the category of songs he wishes to listen to using predefined hand gestures. Finally, after each listening session, the recommendation system reorders the songs in the database so that the songs that the user frequently listens to is at the top of the playlist while the song the user skips are placed at the bottom

KEYWORDS: openCV, MobileNetV2, Haar Cascade, Mediapipe, fisherface.

I. INTRODUCTION

The main objective of our project is to develop an intelligent system that can easily recognize facial expressions and accordingly play a music track based on that particular expression/emotion recognized. Furthermore, we incorporate a hand gesture mechanism for pause/play, increase volume, decrease volume, and skip songs. We have utilized the open-source OpenCV library and Google's Mediapipie API to achieve both tasks. Firstly the music system will detect the face of the user, after that the detected faces will be classified into different emotions based on facial expressions. Then the music player will play music based on the classified emotion. Moreover, the system has a hand gestures recognition system to control the music player, thereby our system does not require any inputs through keyboard, mouse, or touch screens. Finally, the recommendation system personalizes the music player for the user's music preferences. Furthermore, the face mask detection module ensures that if the user is wearing a face mask then he can play songs from any category using pre-mapped hand gestures similar to how happy, angry, sad, or neutral music would have played through facial expression detection.

II. LITERATURE SURVEY

This section presents the existing methods and relevant approaches which are surveyed as follows.

Vasanthakumari, M. U. Peerzade, S. Nayak, Bhagya V, S. Padmanabhan [1] proposed face mask detection using MobileNetV2 with an accuracy of 96%.

F. Zhang, V. Bazarevsky, A. Vakunov, A.Tkachenka, G. Sung, Chuo-Ling Chang, M. Grundmann, [2] proposed mediapie hands, a real-time on-device hand tracking solution that predicts a hand skeleton of a human from a singleRGB camera for AR/VR applications.

P. Hosen, N. Himel, A. Adil, N. N. Moon, F. N. Nur [3] proposed Facial expression recognition for the classification of facial features such as happiness, sadness, fear, disgust, surprise, and anger. A playlist is created based on emotion and user interest and a model is built for the changing wallpaper.

A. Arora, A. Kaul, V. Mittal [4] depicted various algorithms based on classification to provide a clear methodology to i) classify songs into 4 mood categories and ii) detect users mood through his facial expressions and then combine the two to generate user customized music playlist.

L. Y. Mano, B. S. Faical, V. P. Goncalves, G. Pessin [5] proposed an intelligent agent that sorts a music collection based on the emotions conveyed by each song, and then suggests an appropriate playlist to the user based on his/her

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

|| Volume 9, Issue 7, July 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0907135 |

current mood. The user's local music collection is initially clustered based on the emotion the song conveys, i.e. the mood of the song. This is calculated taking into consideration the lyrics of the song, as well as the melody. Every time the user wishes to generate a mood-based playlist, the user takes a picture of themselves at that instant

P. Sarda, S. Halasawade, A. Padmawar [6] proposed a system where the mood is statistically inferred from various data sources primarily: audio, image, text, and sensors. Human's mood is identified from facial expression and speech tones. Physical activities can be detected by sensors that humans usually carry in the form of cellphones. It is a smart music player that keeps learning your listening habits and plays the song preferred by your past habits and mood, activities, etc.

K. Chankuptarat, R. Sriwatanaworachai and S. Chotipant [7] proposed a mobile music player application which is able to recommend songs based on the user's emotion. To classify the user emotion, the proposed application applies both the heart rate and face image.

S. Deebika, K.A. Indira and Dr. Jesline. [8] proposed a mobile music player application which is able to recommend songs based on the user's emotion. To classify the user emotion, the proposed application applies both the heart rate and face image.

A. Sen, D. Popat, H. Shah, P. Kuwor and E. Johri [9], proposed a user-intuitive smartmusic player. This player captures the facial expressions of a person working on the computer and identifies the current emotion. Since various sorts of music are available to boost one's enthusiasm, taking into consideration the tasks executed on the system by the user and the current emotions they carry, an ideal playlist of songs will be created and played for the person.

S. Tiwari and Dr. Aju D [10] proposed a system to detect facial expressions so as to operate an alert system that is very much needed to reduce the crime rate.

O. R. Chanu, A. Pillai, S. Sinha, P. Das [11] proposal illustrates two different techniques of vision-based hand gesture recognition and one data glove based technique. The vision - based technique is static hand gesture recognition technique.

R. Ramanathan, R. Kumaran, R. Rohan R, R. Gupta and V. Prabhu [12] proposed an intelligent agent that sorts a music collection based on the emotions conveyed by each song, and then suggests an appropriate playlist to the user based on his/her current mood. The user's local music collection is initially clustered based on the emotion the song conveys, i.e. the mood of the song

S. Gilda, H. Zafar, C. Soni and K. Waghurdekar [13] presented an effective cross-platform music player, EMP, which recommends music based on the real-time mood of the user. EMP provides smart mood based music recommendation by incorporating the capabilities of emotion context reasoning within our adaptive music recommendation system.

K. S. Nathan, M. Arun and M. S. Kannan [14] proposed EMOSIC, an efficient and accurate model that would generate a playlist based on the current emotional state and behaviour of the user. This proposed system is 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.

A. Howard, C. Zang and E. Horvitz [15] examined the bias issue found in learning algorithms for intelligent systems by focusing on the emotion recognition problem. Firstly baseline outcomes are presented for a cloud-based emotion recognition algorithm applied to images associated with a minority class, in this instance, children's facial expressions. Then a hierarchical approach is presented that combines outputs from the cloud-based emotion recognition algorithm with a specialized learner, and shows that this methodology can increase overall recognition results by 17.3%.

S. G. Kamble and A.H. Kulkarni [16] proposed the detection of a person's expressions by extracting the facial features using the PCA algorithm and Euclidean Distance classifier. The results show that the proposed system achieves up to 84.82% of accuracy level in recognizing the expressions.

III. SYSTEM ANALYSIS AND DESIGN

A. Existing system and their drawbacks

To develop an appropriate application, the existing music player applications are investigated. The drawbacks of the current applications in the market are:

1. The existing systems perform facial expression recognition either on still images or music is played only for the first identified emotion, these systems do not account for emotion changes in real time.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

|| Volume 9, Issue 7, July 2021 ||

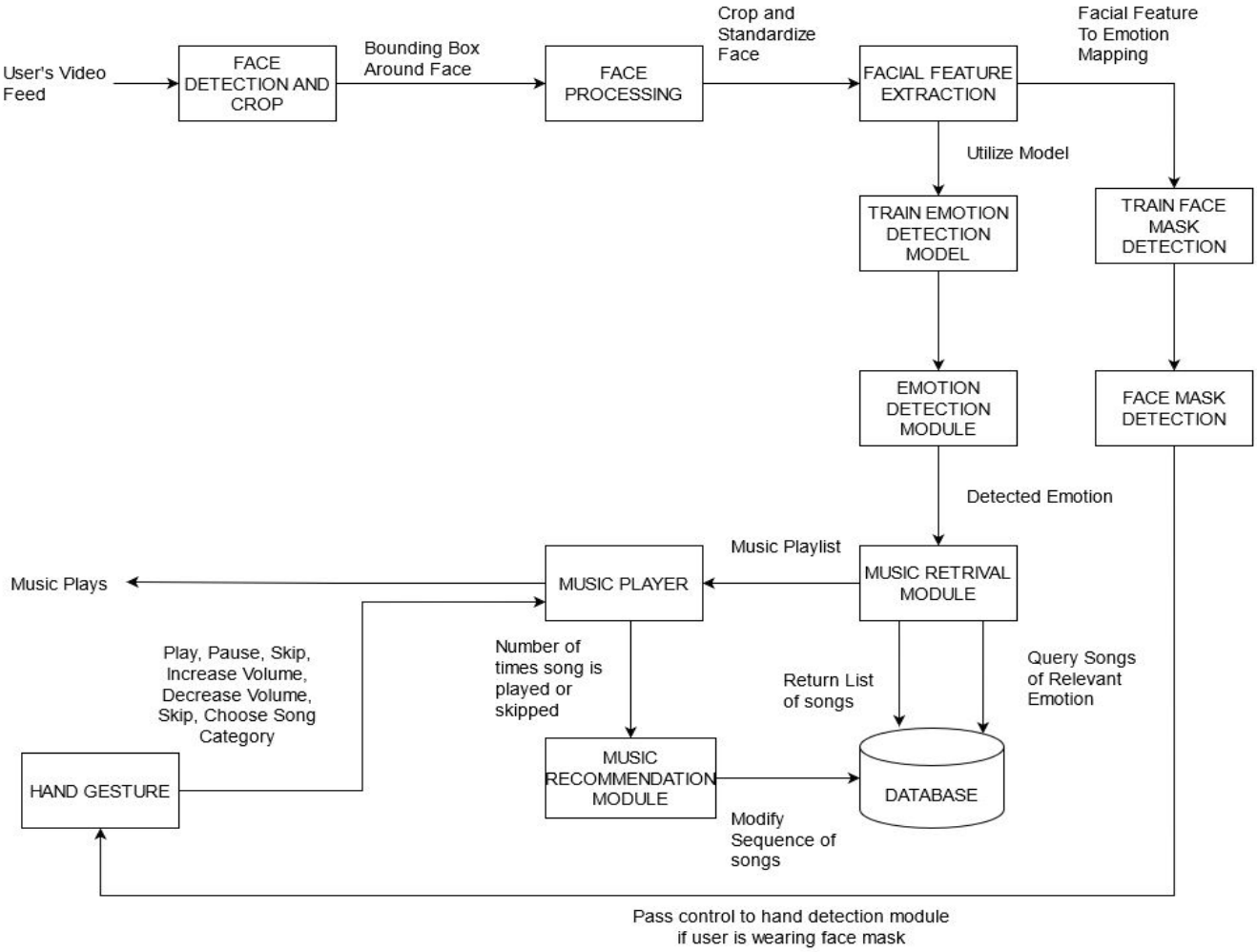
| DOI: 10.15680/IJIRCCE.2021.0907135 |

2. Music control mechanisms require touch interactions with the device.
3. A recommendation system does not exists to understand users preferences

B. System Design

In this section, we will discuss the Architectural System design of our proposed system and Description of the various modules involved.Module and their descriptions are listed below:

1. *Face Detection Module-* This module detects the presence of the user, crops their face, converts to grayscale and saves the image to the dataset. We use the HAAR classifier built into openCV to perform this operation.
2. *Training and Classification Module-* We use images to train a generalized model for emotion detection. We split the obtained dataset into training and classification sets.
3. *Emotion Detection Module-* Our customized model is now deployed to identify the emotion of the user through the real time camera feed. The user's facial expression is determined by the emotion detection module.



4. *Hands Free Control Module-* This module monitors the video feed for predefined hand gestures and on recognition sends pause, play or skip signals to the music player.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

|| Volume 9, Issue 7, July 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0907135 |

5. *Music Retrieval Module*- This module queries the database for the playlist of songs pertaining to the identified emotion. The obtained playlist of songs is forwarded to the Music Player.
6. *Music Player*- The selected music is played on the default music player on the user's computer.
7. *Face Mask Detection Module*- Here we detect if the user is wearing a fast mask. If yes we pass the control to the hand gesture module to specify the category of song to be played.
8. *Music Recommendation Module*- At the end of each session this module must extract statistics from Music player and Hands Free Control Module.

IV. ALGORITHMS

As we have discussed earlier our system includes various modules and functionalities. For the efficient working of the system, a tight integration is required between the different functional modules. This section presents the algorithms used within each module to achieve the respective specific subtask.

A. Face Detection: Haar cascade

For detection of faces in a video feed we use Haar cascade object detection. This method trains a cascade function by applying machine learning on a large set of positive and negative images. Once trained the classifier can detect objects in other images. Through the classifier we can extract features of the face such as nose, eyes etc. But since there are a huge number of potential features we need to extract, to decide the most useful technology known as Adaboost. OpenCV library provides a training method) or pre-trained model for this purpose. Input of the algorithm is a HAAR classifier supplied with OpenCV called haarcascade_frontalface_default. Output includes Detected Users Face since all human faces share some similar properties.

B. Emotion Recognition: Fisher Face

Fisherface is one of the popular algorithms used in face recognition, and is widely believed to be superior to other techniques, such as eigenface because of the effort to maximize the separation between classes in the training process. The Image recognition using fisherface method is based on the reduction of face space dimension using Principal Component Analysis (PCA) method, then applying Fisher's Linear Discriminant (FDL) method or also known as Linear Discriminant Analysis (LDA) method to obtain features of image characteristic. The algorithm used in the process for image recognition is fisherfaces algorithm while for identification or matching face image using minimum euclidean methods. OpenCV provides built-in functionality to create and apply Fisher Face Objects In our project we have classified facial expressions into 4 categories - happy, angry, sad and neutral.

C. Hand Gestures for Music Control: Mediapipe Hands

To achieve the music control functionality through the music player we detect the finger or combination of fingers held upright by the user's left hand. We then map each combination to a particular control signal. ex:- index finger raised is mapped to pause and play, index finger raised along with middle finger signifies increased volume and so on. The actual implementation of the system is achieved through the MediaPipe ML Library developed by google. MediaPipe Hands is a high-fidelity hand and finger tracking solution. It employs machine learning (ML) to infer 21 3D landmarks of a hand from just a single frame. It utilizes an ML pipeline consisting of multiple models working together: A palm detection model that operates on the full image and returns an oriented hand bounding box. A hand landmark model that operates on the cropped image region defined by the palm detector and returns high-fidelity 3D hand keypoints.

D. Training Technique for Facial Expression Model Personalization

To train the facial expression detection model. We instruct the user to depict a particular emotion, say happy and take multiple images, preprocess it and save it. We do this for all 4 emotions (Happy, Angry, Sad and Neutral). Using the images we train the detection model using openCV and fisherfaces.

E. Face mask detection: MobileNetV2

We use MobileNetV2 to train the face mask detection model. Then use the same model to determine if the user is wearing a face mask or not. If the user is wearing a mask then we transfer the control to the hand gesture module and instruct the user to specify the category of music using a pre mapped hand gesture.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

|| Volume 9, Issue 7, July 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0907135 |

Step 1: Image Frame is extracted from the webcam feed and provided to the system after preprocessing

[i] Image data to array

[ii] Label Binarization

[iii] Split Dataset into Training and Testing

Step 2: The input is sent through multiple layers of MobileNetV2

[i] Set Slow Learning Rate, Epoch and Batch Size

[ii] Image Augmentation

Step 3: Pooling using 7x7 kernels.

Step 4: We perform ReLU activation with 128 dense layers

Step 5: Dropout to prevent overfitting

Step 6: Finally we apply softmax activation with 2 dense channels

Step 7: We can now detect if the video frame ROI has a mask or not

F. Music Recommendation Algorithm

Step 1: Retrieve Number of times song has been played

Step 2: If song has never been played

Do: Favorability Score Remains the Same

Step 3. If song has been played more than n times:

Do: Increase Favorability Score

Step 4. Number of time song has been skipped:

Do: Decrease Favorability Score

Step 5. Reorder songs in the database in decreasing order of favorability score

V. APPLICATIONS

The applications of the proposed system are as follows:

1. Can be used in automobiles where the users can focus on driving and music is played to them automatically.
2. Can be used in risk prone workspaces like kitchen, garage and workshops due to hands free music control.
3. Can be used in smart TV's at home to play content even without touching the remote.
4. Can be integrated into smart music apps such as spotify or youtube to serve more personalized content.
5. A more accessible Music player useful for differently abled users.

VI. CONCLUSION AND FUTURE WORK

The Emotion-Based Music Player is used to automate music discovery and to give a better music player experience for the end user. The application solves the basic needs of music listeners without disturbing their workflow. The gesture control system improves the accessibility of the music player system. Finally the face mask detection makes the music player usable in the current pandemic situation.

Future implementation would be in connecting our application to music streaming services such as Youtube music or Spotify to discover and play any song the user desires.

REFERENCES

[1] Vasanthakumari, M. U. Peerzade, S. Nayak, Bhagya V, S. Padmanabhan, "Automatic Facemask Detection using MobileNetV2", *International Research Journal of Engineering and Technology(IRJET),* Feb 2021.

[2] F. Zhang, V. Bazarevsky, A. Vakunov, A.Tkachenka, G. Sung, Chuo-Ling Chang, M. Grundmann, "MediaPipe Hands: On-device Real-time Hand Tracking", *Google Research*, 2020.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

|| Volume 9, Issue 7, July 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0907135 |

[3] P. Hosen, N. Himel, A. Adil, N. N. Moon, F. N. Nur, "*Music Playing and Wallpaper Changing System Based on Emotion from Facial Expression*," Advances in Intelligent Systems and Computing, December 2019.
[4] A. Arora, A. Kaul, V. Mittal, "*Mood Based Music Player*," Springer Soft Computing, 2019.
[5] L. Y. Mano, B. S. Faical, V. P. Goncalves, G. Pessin, "*An Intelligent and generic approach for detecting human emotions: a case study with facial expressions*" IEEE 5th Int. Conf. for Convergence in Technology, 2019.
[6] P. Sarda, S. Halasawade, A. Padmawar, "*Emousic: Emotion and Activity Based Music Player Using Machine Learning*," Springer Advances in Computer Communication and Computational Sciences, May 2019.
[7] K. Chankuptarat, R. Sriwatanaworachai and S. Chotipant, "*Emotion-Based Music Player*," IEEE 5th Int. Conf. on Engineering, Applied Sciences and Technology (ICEAST), 2019.
[8] S. Deebika, K.A. Indira and Dr. Jesline, "*A Machine Learning Based Music Player By Detecting Emotions*," Fifth International Conference on Science Technology Engineering and Mathematics, 2019.
[9] A. Sen, D. Popat, H. Shah, P. Kuwor and E. Johri, "*Music Playlist Generation Using Facial Expression Analysis and Task Extraction*," Int. Comm. And Computational Tech. Springer, 2018.
[10] S. Tiwari and Dr. Aju D, "*Operating an Alert System using Facial Expression*," Int. Conf. on Innovations in Power and Advanced Computing Tech., 2017.
[11] O. R. Chanu, A. Pillai, S. Sinha, P. Das, "*Comparative Study For Vision Based and Data Based Hand Gesture Recognition Technique*," IEEE Int. Conf. on Int. Comm. and Computational Technique, 2017.
[12] R. Ramanathan, R. Kumaran, R. Rohan R, R. Gupta and V. Prabhu, "*An Intelligent Music Player based on Emotion Recognition*," 2nd IEEE Int. Conf. on Comput. Syst. and Inf. Tech. for Sustain. Solutions, 2017.
[13] S. Gilda, H. Zafar, C. Soni and K. Waghurdekar, "*Smart Music Player Integrating Facial Emotion Recognition and Music Mood Recommendation*," IEEE WiSPNET, 2017.
[14] K. S. Nathan, M. Arun and M. S. Kannan, "*EMOSIC — An emotion based music player for Android*," in IEEE Int. Symp. on Signal Processing and Information Technology, Bilbao, 2017.
[15] A. Howard, C. Zang and E. Horvitz, "*Addressing Bias in Machine Learning Algorithms: A Pilot Study on Emotion Recognition for Intelligent Systems*", Microsoft Research, 2017.
[16] S. G. Kamble and A.H. Kulkarni, "*Facial Expression Based Music Player*", Intl. Conf. on Advances in Computing, Comm. And Info, 2017.

INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 7.542

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details