International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 1, Issue 3, pp 40-44, December-2017

Implementation of the Audio Feature Extraction Module System from the Detection in the Human Emotions

G Naga Kumar Kakarla

Associate Professor, Department of Computer Science and Engineering, Malla Reddy College of Engineering for Women, Hyderabad -500100, Telangana, India

***Abstract* - The mood of an individual person is usually recognized based on their facial expressions. With today's technologies, distinguishable features of the face can be extracted as inputs with the help of a webcam or any other external device. The gathered data helps in detecting the mood and songs are played from a personalized playlist, if available or a default playlist can be used based on the mood detected. This removes the time-consuming and tedious task of manually grouping songs into different lists and helps in generating an appropriate playlist based on an individual's emotional features. Thus, our proposed system mainly aims on detecting human emotions for developing emotion-based music player. A brief idea about our systems working, playlist generation and emotion classification is mentioned below.**

Keywords: Emotion Recognition; music recommendation; Facial Extraction; Emotion Extraction Module; Audio Feature Extraction Module.

1. INTRODUCTION

Music listeners have a tough time creating and segregating the playlist manually when they have hundreds of songs. It is also difficult to keep track of all the songs: sometimes songs that are added and never used, wasting a lot of device memory and forcing the user to find and delete songs manually. Users have to manually select songs every time based on interest and mood. User's also have difficulty to re-organize and playing music when play-style varies. So, we have used Machine Learning concept which involves facial scanning and feature tracking to determine the user's mood and based on it gives a personalized playlist.

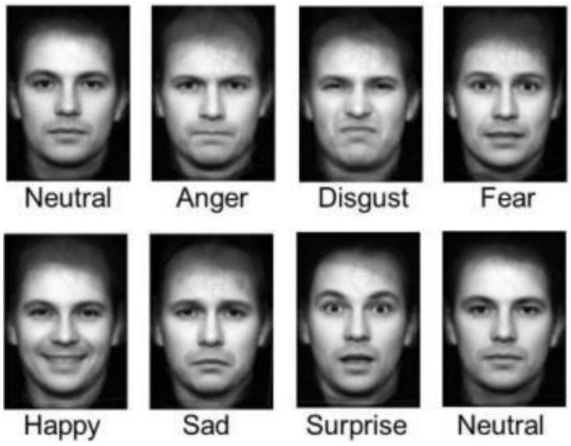


Fig -1: Various basic emotions of humans

2. LITERATURE SURVEY

S Metilda Florence and M Uma (2020) proposed a paper "Emotional Detection and Music Recommendation System based on User Facial Expression" where the proposed system can detect the facial expressions of the user and based on his/her facial expressions extract the facial landmarks, which would then be classified to get a particular emotion of the user. Once the emotion has been classified the songs matching the user's emotions would be shown to the user. It could assist a user to make a decision regarding which music one should listen to helping the user to reduce his/her stress levels. The user would not have to waste any time in searching or to look up for songs. The proposed architecture contained three modules, namely, Emotion extraction module, Audio extraction module and Emotion-Audio extraction module. Although it had some limitations like the proposed system was not able to record all the emotions correctly due to the less availability of the images in the image dataset being used. The image that is fed into the classifier should be taken in a well-lit atmosphere for the classifier to give accurate results. The quality of the image should be at least higher than 320p for the classifier to predict the emotion of the user accurately. Handcrafted features often lack enough generalizability in the wild settings.

H. Immanuel James, J. James Anto Arnold, J. Maria Masilla Ruban, M. Tamilarasan (2019) proposed "Emotion Based Music Recommendation" which aims at scanning and interpreting the facial emotions and creating a playlist accordingly. The tedious task of manually Segregating or grouping songs into different lists is reduced by generating an appropriate playlist based on an individual's emotional features. The proposed system focuses on detecting human emotions for developing emotion-based music players. Linear classifier is used for face detection. A facial landmark map of a given face image is created based on the pixel's intensity values indexed of each point using regression trees trained with a gradient boosting algorithm. A multiclass SVM Classifier is used to classify emotions Emotions are classified as Happy, Angry, Sad or Surprise. The limitations are that the proposed system is still not able to record all the emotions correctly due to the less availability of the images in the image dataset being used. Diverse emotions are not found. Handcrafted features often lack enough generalizability in the wild settings.

Ali Mollahosseini, Behzad Hasani and Mohammad H. Mahoor (2017) proposed "AffectNet: A Database for Facial Expression, Valence, and Arousal Computing in the Wild" where more than 1,000,000 facial images were obtained from the Internet by querying three major search engines using1250 emotion related keywords in six different languages. About half of the retrieved images were manually annotated for the presence of seven discrete facial expressions and the intensity of valence and arousal. Two baselines are proposed to classify images in the categorical model and predict the value of valence and arousal in the continuous domain of dimensional model. There were certain limitations such that VGG16 only makes improvement over AlexNet by replacing large kernel sized filters with multiple 3X3 kernel-sized filters one after another. With a given receptive field multiple stacked smaller size kernel is better than the one with a larger size kernel. AffectNet database does not contain very strong samples. That is, samples where valence is 1 or -1 and with arousal being 1 or -1.

3. PROPOSED SYSTEM

Our approach is to use Deep Neural Networks (DNN) to learn the most appropriate feature abstractions directly from the data taken in an uncontrolled environment and handle the limitations of handcrafted features. DNNs have been a recent successful approach in visual object recognition, human pose estimation, face verification and many more. Availability of computing power and existing big databases allow DNNs to extract highly discriminative features from the data samples. CNNs are proven to be very effective in areas such as image recognition and classification. CNNs are very effective in reducing the number of parameters without losing on the quality of models.

The proposed system can detect the facial expressions of the user and based on individual's facial expressions using VGG16 CNN model. Once the emotion has been classified the song matching the user's emotions would be played.

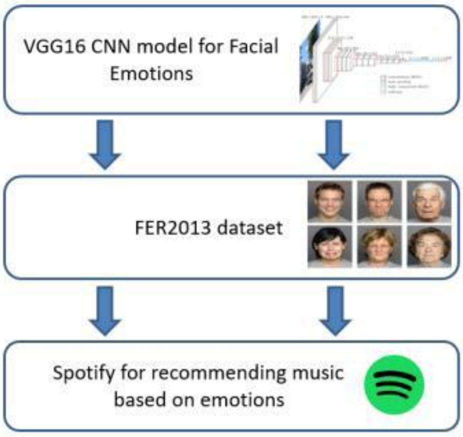
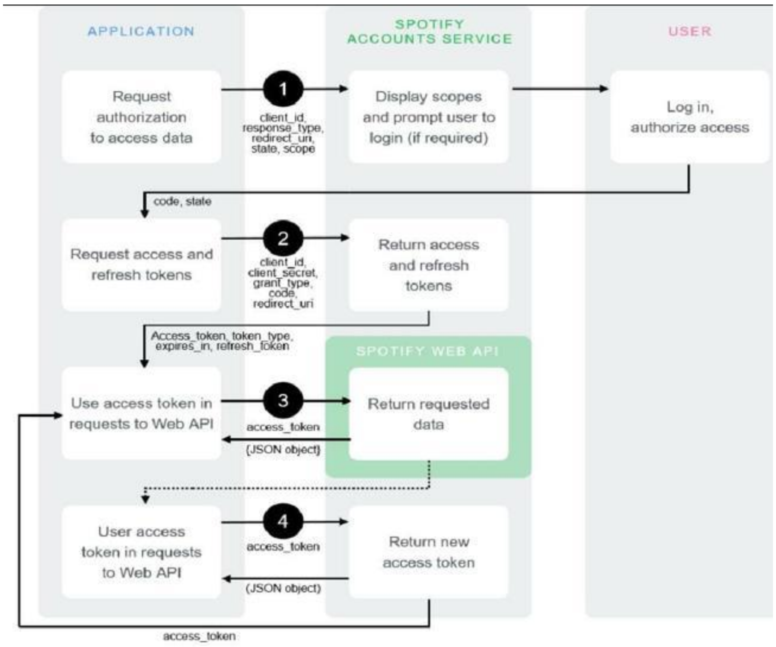

VGG16 CNN model for Facial Emotions
FER2013 dataset
Spotify for recommending music based on emotions

Fig -2: Proposed System

4. SYSTEM ARCHITECTURE

In this project, a main webpage is designed where a user has to record his/her video stream by clicking on the start video stream button present on the webpage. To detect his/her mood, he/she has to click the capture button and the image is captured by the client side and sent to the server side where the captured image is being processed by the VGG 16 model and prediction is made. The corresponding prediction made is matched with the emotion labels and the emotion will be displayed on the webpage. Currently there are 7 emotion labels i.e., Happy, Sad, Neutral, Disgust, Angry, Fear and Surprised. Once the emotion is detected, the next phase is to play songs based on the mood. We have integrated our webpage with Spotify API. The Client side then requests tracks from the Spotify API and the songs and audio features are analyzed and songs are played to the attending user based on the emotion detected. Here the playlist can be either given by the user and from that the songs will be played based on the detected emotion or a default playlist is used. The user is also given Play and Pause option and can use it whenever needed.

The system architecture consists of five modules namely client, user, server, VGG16 model and Spotify API. The client program captures an image of the user and sends it to the server. The server takes the image, greyscale it and checks whether a face can be found using the haar cascade algorithm. Face found in the image is cropped and sent to the pretrained VGG16 model. Based on the predictions returned by the model, the server labels an emotion for the image and sends it to the client. The client requests songs and its audio features in the user's playlist from the Spotify API and analyses the audio features. Based on the analysis, the client suggests a suitable song to the user.


APPLICATION
SPOTIFY ACCOUNTS SERVICE
USER
Request authorization to access data
1
client_id, response_type, redirect_uri, state, scope
Display scopes and prompt user to login (if required)
Log in, authorize access
code, state
Request access and refresh tokens
2
client_id, client_secret, grant_type, code, redirect_uri
Return access and refresh tokens
Access_token, token_type, expires_in, refresh_token
SPOTIFY WEB API
Use access token in requests to Web API
3
access_token
(JSON object)
Return requested data
User access token in requests to Web API
4
access_token
(JSON object)
Return new access token
access_token

5. SYSTEM IMPLEMENTATION

Emotion Detection Implementation

The emotion function takes care of emotion detection. The pretrained model is loaded using the load_model function in TensorFlow. HaarCascade classifier is also loaded from the OpenCV package. The emotion function takes an image as input. It pre-processes the image and applies HaarCascade classifier. If no face is found nil label is returned else it takes the last face found and sends it to the loaded vgg16 model. The model makes predictions about the image. The predictions are analysed and emotion label is returned.

Spotify Implementation

The requestAuthorization() function takes care of the sets the parameters for requesting authorization and procuring an access token. The callAuthorizationApi() function sends a post request to the Spotify accounts service which returns the access token. This token is used in every request that are sent here after for authentication purposes. The callApi() is used to send requests to the Spotify accounts service for information like tracks in a playlist, all the devices that are currently active, audio features that available for a particular track etc.

Server Implementation

The server is built on a flask framework. It has two routes namely home and emotion. The home route returns the application webpage and the emotion route receives an image and returns an emotion label. The emotion route uses the emotion function for emotion detection.

6. RESULTS ANALYSIS

The VGG16 model has been trained for over 400 epochs with 478 images in each batch. Throughout the course of training various hyperparameters are tuned. The Google Collab platform is used for training. Accuracy graphs captured during training are plotted below.

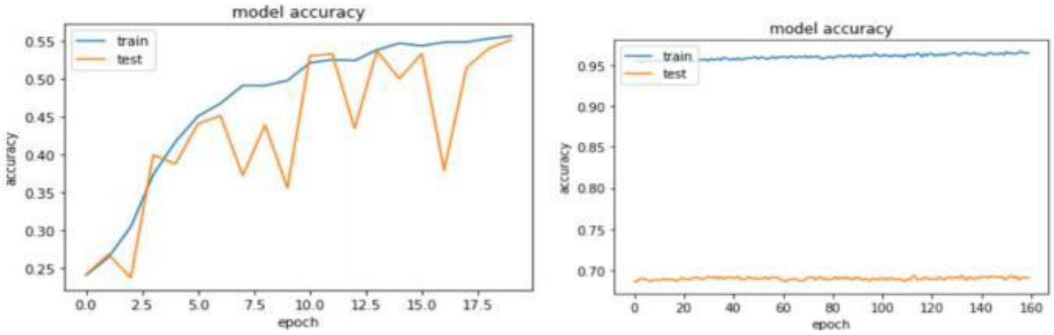
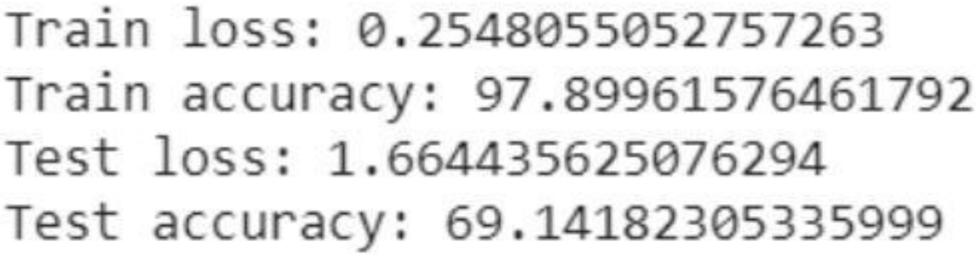


Fig -4: The left snapshot is Epoch 1 – 20 where learning_rate=0.01, momentum=0.95 ,epochs=20 and the right snapshot is Epoch 160 – 240 where learning_rate=0.00001, momentum=0.95 ,epochs=80.

Our VGG16 model achieved an accuracy of around 98% in Train set and around 70% in Test set.



```
Train loss: 0.2548055052757263
Train accuracy: 97.89961576461792
Test loss: 1.664435625076294
Test accuracy: 69.14182305335999
```

Fig -5: Final test & train accuracy

Below are some screenshots taken from the application where the user has to start the video stream and capture the image. The emotion is detected and songs are played from the Spotify playlist.

7. CONCLUSION AND FUTURE WORK

The application aims to provide a simpler, additional hardware-free and reliable emotion-based music system to the operating system users. The Emotion-based music program would help people who are searching for music driven on the emotion and emotional behaviour. It could help to reduce the search time for music and thus reduce the unnecessary computational time and thus increase the overall accuracy and efficiency of the system. The application solves the basic needs of music listeners without troubling them as existing applications do: it uses technology to increase the interaction of the system with the user in many ways. It eases the work of the end-user by capturing the image using a camera, determining their emotion, and suggesting a customized playlist from Spotify Premium account through a more advanced and interactive system.

- Facial recognition can be used for authentication purposes.
- Could be implemented in Raspberry Pi as a feature of Smart Home.
- Can be used to determine the mood of physically challenged & mentally challenged people.
- Music classifier models can be developed.

REFERENCES

[1] Metilda Florence S and Uma M, 2020, “Emotional Detection and Music Recommendation System based on User Facial Expression”, IOP Conf. Ser.: Mater. Sci. Eng. 912,06/2007.

[2] EMOTION BASED MUSIC RECOMMENDATION SYSTEM H. Immanuel James, J. James Anto Arnold, J. Maria Masilla Ruban, M. Tamilarasan, R. Saranya IRJET (2019)

[3] Ali Mollahosseini, Behzad Hasani and Mohammad H. Mahoor 2017, “AffectNet: A Database for Facial Expression, Valence, and Arousal Computing in the Wild”, arXiv:1708.03985v4[cs.CV],10/2017

[4] Emophony – Face Emotion Based Music Player Banpreet Singh Chhabra – IRJET (JUNE 2020)

[5] Seungjae Lee, Jung Hyun Kim, Sung Min Kim, & Won Young Yoo. (2011). Smoodi: Mood-based music recommendation player, 2011 IEEE International Conference on Multimedia and Expo.