Music Recommendation using Emotion Analysis

Anita Chaudhari[1], Anuj Patil[2], Mahesh Tiwari[3], Deepanshu Suthar[4]
Department of Information Technology, St. John College of Engineering and Management, Palghar, India

Abstract:- In this era of internet, all the music is now available on the commercial music services in abundance. The quantity of the available music pieces is in dozens of millions. This makes choosing an appropriate music piece out of this abundance of music is a time-consuming task. Whenever users want to listen to music of specific genre, a need to find a music piece similar to what user likes to hear arises. Hence it is very important to develop a system that can select music to a user based on certain criteria. This system is designed to recommend music to users according to their mood by recognizing their facial expressions using machine learning algorithm to decide a user's emotional state.

Keywords:- *OpenCV, Facial Emotion Recognition, Haarcascade, Music Recommendation etc.*

I. INTRODUCTION

The property of music to trigger emotions in humans is the main reason of widespread use of music in daily lives. Most people are aware about the ability of music triggering emotions such as happy, sad, and angry etc. But content-based methods recommend music pieces that are similar to user's favourites. But in actual scenario, users don't rate every music piece they like. Rather a very small amount of music pieces is actually rated by users which makes the pool of user reviews smaller. In turn, this makes the system recommend a very narrow range of music pieces to the user. Because a lot of focus is given on developing music retrieval systems in which user preference representing queries are prepared. Music recommendation is needed today is because we can access a huge number of music pieces on the internet. To find songs using retrieval systems, we have to execute queries manually and repeatedly and hence it is difficult to find about which queries are appropriate. To overcome this, it is expected that a recommender system selects probably-preferred pieces from the database by estimating user preferences.

The two most widely used classification techniques are collaborative filtering approach and content-based filtering approach.

1. Collaborative methods recommend pieces to a user by taking ratings of those pieces by other people. For example, consider there are three users. User is a person who likes music pieces X and Y. If there are many other users who like the music pieces X, Y, and Z then the music Z will most likely be recommended that user. This method is widely used in E-commerce services (e.g., Amazon and Flipkart) and is proven to be effective. But this system doesn't work well for new content, i.e. music that is recently released or hasn't been played too many times.
2. Content-based filtering recommends music which is similar to a user 's favourites music in terms of its features. This leads to a great artist variety; with various music pieces including unrated ones being recommended.

II. MATERIAL AND METHODS

i. LITERATURE REVIEW

In 2009, Byeong-jun Han *et al.*[1] presented a content-based recommendation system. The paper aimed at developing a hybrid system that used both content-based recommendation with music ontology. A model was trained by analysing low-level features of music. It was proposed in three steps. An emotion state transition model (ESTM), an emotion transition matrix (ETM) and an ontology to depict emotions and a model was proposed to map the emotion transition with low level features of music.

Kyoungro Yoon *et al.* [2] proposed developing a system that used collaborative approach for music recommendation. For this, it uses low level features of music pieces that can trigger an emotional response from an individual. The system used selected low-level music features tackle scalability problem that occurs in tag-based system. The system was based on 6 high-level features, 28 mid-level features and 890 low-level features. The drawbacks that accompanied by either content-based or collaborative approach was addressed by Marius Kaminskas *et al.* [3] Content-based systems study features like genre, artist etc. to find similarity. This method tackles the problem of popularity bias in collaborative filtering. But systems that use audio features need high computational power. Modern commercial music libraries are humongous in sizes, and hence useless for commercial use. Both filtering methods have drawbacks and hence there have been many attempts to combine both the approaches. Hence authors [4] proposed a hybrid system that combines both the filtering algorithms. For music similarity, they used waited Euclidean algorithm using users listening history which addresses both the filtering techniques.

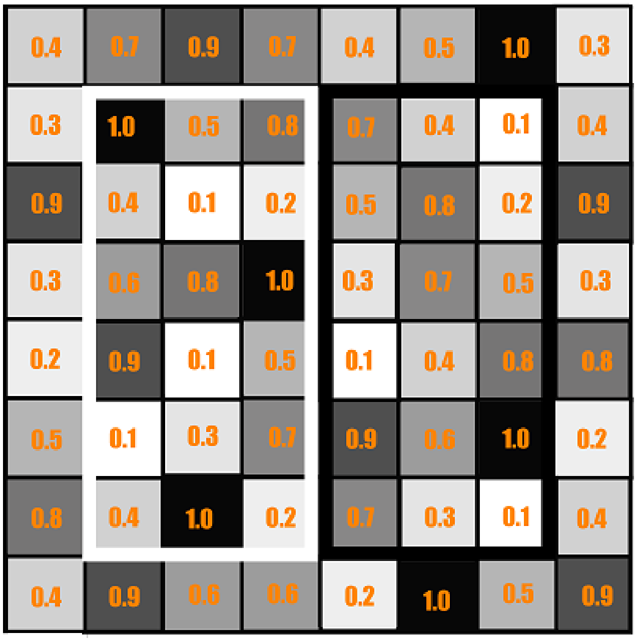
ii. HAAR CASCADE CLASSIFIERS

Object Detection is used to detect various shapes, objects and people from image or video feeds. This method was proposed by authors [5]. The basis of Haar object classifier and detection method is the Haar-like features. Instead of using the intensity of a pixel, haar features use variance in contrast between neighbouring pixel groups for detection of object's edge. The difference between the contrast of various pixel groups is used to determine light and dark areas in an image. A group of points with contrast variance makes up a haar-like feature. Haar-like features are used to detect an object or person within an image. Haar features are scalable by increasing or decreasing the pixel group size being examined.

The features of an image are calculated using the integral image. The integral image is an array of sums of the intensity values in a pixel group. A pixel group is formed by selecting a pixel above the subject pixel and a pixel left to the subject pixel.

iii. HAAR FEATURES

Haar features are used to find the object's edges or to identify the pixel groups with high intensity variances.



0.4	0.7	0.9	0.7	0.4	0.5	1.0	0.3
0.3	1.0	0.5	0.8	0.7	0.4	0.1	0.4
0.9	0.4	0.1	0.2	0.5	0.8	0.2	0.9
0.3	0.6	0.8	1.0	0.3	0.7	0.5	0.3
0.2	0.9	0.1	0.5	0.1	0.4	0.8	0.8
0.5	0.1	0.3	0.7	0.9	0.6	1.0	0.2
0.8	0.4	1.0	0.2	0.7	0.3	0.1	0.4
0.4	0.9	0.6	0.6	0.2	1.0	0.5	0.9

Fig 1. Representation of an image using haar features [6]

The fig. 1 represents an image with pixel group values ranging from 0 to 1. The haar-features are calculated by finding the difference between the average pixel intensities in darker and brighter region. It is considered an edge when the difference is 1. The haar features are calculated by traversing in a raster scan approach pixel by pixel. Haar feature calculation requires a lot of calculations. Calculating all these features is a resource heavy task and hence the concept of integral image was introduced. Sum of pixels in left and above a pixel in the original image forms a single pixel in integral image. Hence the bottom right pixel is sum of all the pixels in the Original Image. Integral images reduce the time complexity greatly.

iv. HAAR FEATURE SELECTION

A large set of features is decided to capture certain structures in an image. The number of these features is 6000. So a feature selection technique is needed to only select the relevant features. A subset of 6000 features is ran against training images. Not all the features are tested at once. Simple features are tested first. If no positives are found, then only next subset of features is considered.

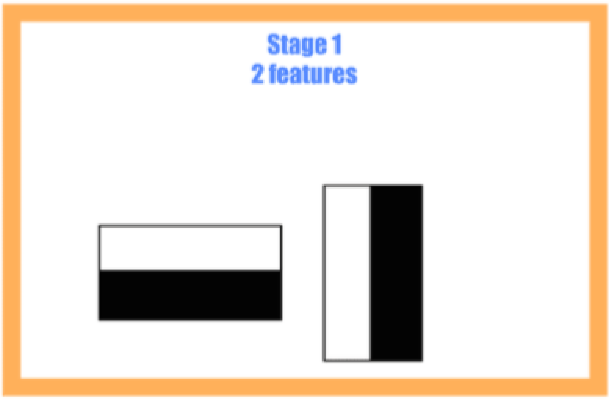


Fig 2. Phase 1 Feature testing [6]

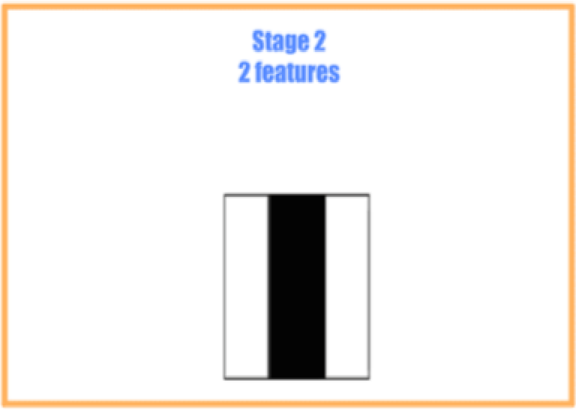


Fig 3. Phase 2 Feature testing. [6]

In first stage, two simple features are selected. For the second stage, a single complex feature is selected. Since the number of features is huge, the workload is reduced using this technique.

v. DATASET

The FER2013 dataset was used for training the model. FER2013 is a large dataset of images which is publicly accessible under Kaggle's FER challenge. It contains images with seven different emotional facial features like anger, disgust, sadness etc.

III. PROPOSED METHODOLOGY

i. PREPROCESSING

Pre-processing is used to enhance the performance of the subsequent stages. Pre-processing and cleaning of data is needed to be done to improve the performance of feature selection and classification.

The FER2013 dataset contains that are already cropped and aligned to the face of a single subject in an image. For training, it is faster and more efficient to train a model on greyscale images instead of a fully coloured image. Haar cascade that can also address and rectify the issue of ambient occlusion, illumination and head pose issues. We use histogram equalization to detect and correct the illumination issues.

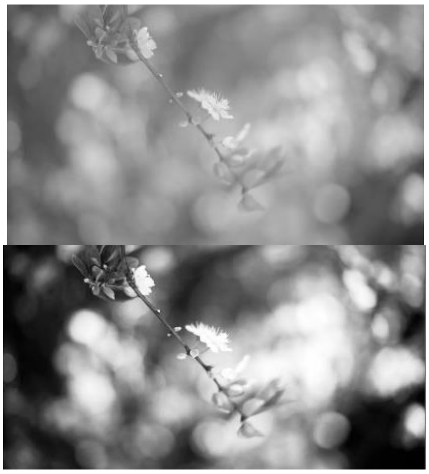


Fig 4. Histogram Equalization [7]

ii. FEATURE EXTRACTION

Extracting facial features means translating the input images into a set of distinct feature. Feature extraction is used to reduce the immense amount of data to a very small manageable size so as to make it more processable. We use haar-cascade classified for extraction of features known as low learners. Images are tested against knowledge of 6000 different facial feature identifiers and extracted when the features are reported positive. To reduce computational load, we use a small subset of features which are proven to define a face.

iii. TRAINING CLASSIFIER

Two sets of images are needed for training the classifier. First set has images that don't any distinct objects or features. These are called negative images. Other set contains positive images which contains more than one set of objects. To train facial features, a dataset of more than 4000 negative images were selected.

To produce an accurate facial feature detection, a original positive set of images is selected. This set contains images of people with different facial feature, gender, age and skin complexion

iv. ARCHITECTURE

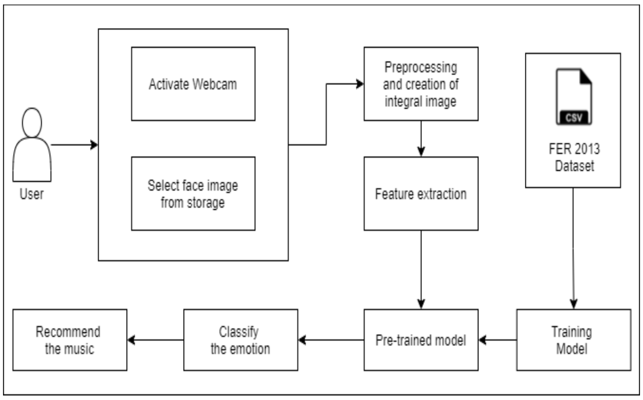


Fig 5. Architecture

The classifier is trained with images from FAR 2013 dataset. In this, a subset of images is stated as positive and remaining images from dataset are negative i.e. without any faces.

The system presents user with a UI. User needs to login to the system as shown in Fig 6. On successful login, user is presented with two options whether to detect emotion in an image from storage or to use webcam for capturing the image. After user selects one of the options, an image frame is captured from the image provided. This image is then converted to a greyscale image and later into an integral image using Haar cascade algorithm. Same algorithm extracts facial features from the image. these features are then compared with the model which was pre-trained using FAR2013 dataset. The emotion is then classified into either of the six defined emotions viz. Happy, angry, sad, excited, neutral. As per user's detected mood, a pre-compiled song playlist is fed to the default media player of the device. User doesn't need a separate UI for music as the default media players have their own navigation and control buttons that the user can use to control the songs.

v. MUSIC RECOMMENDATION

The input is acquired in real-time so the camera is used to capture the video and then the framing is done.

After an emotion is detected, a pre-compiled list of songs related to that emotion is played on a local media player of the device. Since the song playlist is shown in local media player, user can listen to any song he/she would like to and navigate through the playlist. Based on the regularity that the user would listen to songs, the songs are displayed in that order.

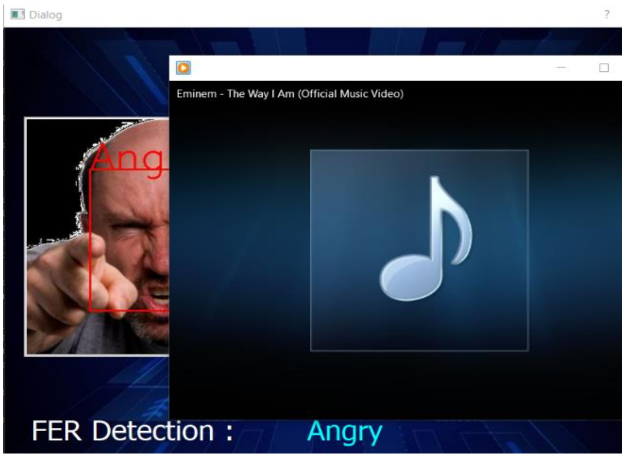


Fig 6. Playing music in local music player

vi. USER INTERFACE

The user interface has a login windows. Either as a user or as an admin. Admin can train or make changes to the model. While user gets two options to either make a choice to detect using an already captured image or to use webcam to detect emotion from live video feed from webcam.

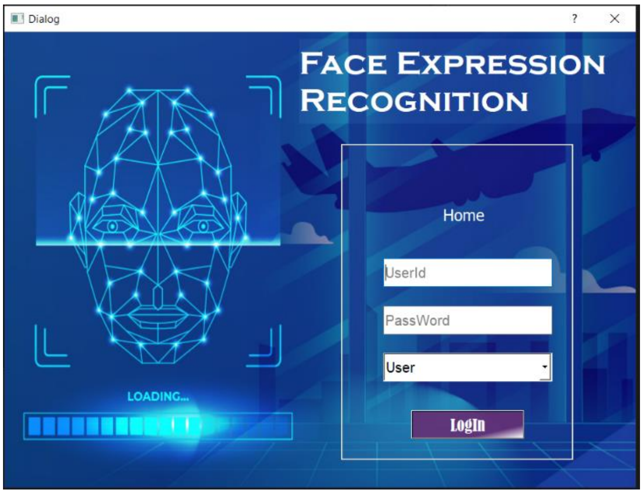


Fig 7. Login screen of the application

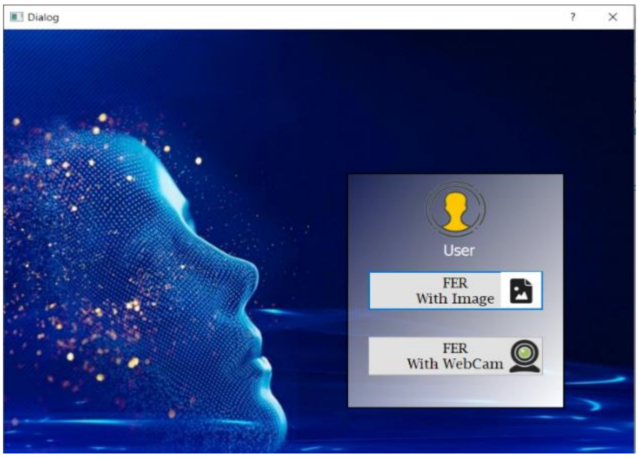


Fig 8. User options

IV. CONCLUSION & FUTURE WORK

In this paper, presented a method to detect facial emotions using Haar-cascade object detection machine learning algorithm. The algorithm detects a large number of distinct facial features that compare images objects based on the variance of the intensity of pixel groups. An additional haar-like feature is introduced in classifier called histogram equalization so as to reduce the inaccuracies introduced by illumination issues. The system then recommends music pieces to user based on the detected emotional state of user, the music list is a precompiled playlist based on content based filtering. The music is played on the local machine of the user and provides navigation controls of the local media player to navigate n control the music.

REFERENCES

[1]. Byeong-jun Han et al. “Music emotion classification and context-based music recommendation”, *Multimed Tools Appl Vol. 47 Issue 3, (2010)*

[2]. Kyoungro Yoon et al. “Music Recommendation System Using Emotion Triggering Low-level Features”, *IEEE Transactions on Consumer Electronics, Vol. 58, No. 2, (2012)*

[3]. Marius Kaminskas et al. “Location-Aware Music Recommendation Using Auto-Tagging and Hybrid Matching”, *Proceedings of ACM conference on Recommender systems, Vol.7, (2013)*

[4]. Cheng-CheLu and Vincent S.Tseng, “A novel method for personalized music recommendation”, *Expert Systems with Applications, Vol. 36, Issue 6, Aug 2009*

[5]. Paul Viola and Michael Jones, “Rapid Object Detection using a Boosted Cascade of Simple Features”, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition 1:I-511- I-518 vol.1, Feb 2001*

[6]. Face Detection with Haar Cascade, https://towardsdatascience.com/face-detection

[7]. A Tutorial to Histogram Equalization, A Tutorial to Histogram Equalization by Kyaw Saw Htoon

[8]. Ziwon Hyung et al. “Music recommendation using text analysis on song requests to radio stations”, *Expert Systems with Applications Vol. 41, (2014)*

[9]. S Metilda Florence , M Uma, “Emotional Detection and Music Recommendation System based on User Facial Expression”, *3rd International Conference on Advances in Mechanical Engineering (ICAME 2020)*

[10]. Shekhar Singh, Fatma Nasoz, “Facial Expression Recognition with Convolutional Neural Networks”, *10th Annual Computing and Communication Workshop and Conference (CCWC), 6-8 Jan. 2020.*

[11]. Zhang K, Zhang Z, Li Z., “Joint face detection and alignment using multitask cascaded convolutional networks”, *IEEE signal Processing Letters, 2016, 23(10):1499-1503*

[12]. Adolf F., “How-to build a cascade of boosted classifiers based on Haar-like features” http://robotik.inflomatik.info/other/opencv, June 20 2003

[13]. Pavate, Aruna & Chaudhari, Anita & Bansode, Rajesh. (2019). Envision of Route Safety Direction Using Machine Learning. Acta Scientific Medical Sciences. 3. 140-145. 10.31080/ASMS.2019.03.0452.