IJARIIT INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X
Impact Factor: 6.078
(Volume 8, Issue 5 - V8I5-1246)
Available online at: https://www.ijariit.com

Song recommendation system using emotion detection

Shubham Chorage
shubham.s.chorage@gmail.com
Pune Institute of Computer Technology, Pune, Maharashtra

ABSTRACT

Human emotion detection is an immediate need so that modern AI systems can simulate and measure facial responses. It also has advantages in identification of intent, promotion of products and security verification. Real time emotion recognition from images and video is a very simple task for the human eyes and brain, but it proves to be very difficult for machines and similar machine learning tools. Basically, Image Processing techniques are needed for feature extraction supported by a reliable database trained in a Machine Learning model for the system. Several machine learning algorithms and tools such as Convolutional Neural Network, OpenCV, Deep Learning, Eigen values and Eigen vectors are suitable for this job and I intend to use these methods in this project. For machines development of different modules and then training them using various images and real time feed is essential. Various leading institutions and researchers have trained their own model for an accuracy of approximately 50% and above. This project explores the ML algorithms as well as emotion detection techniques which would help us in correct identification of the human emotion and furthermore, implementation of the system into useful application of a music recommendation system.

Keywords: *Emotion Recognition, Deep Learning, Music Recommendation System, Human-Computer Interaction, Convolutional Neural Network*

1. INTRODUCTION

The face is our essential focal point of consideration in public activity assuming a significant part in passing on personality and feelings. We can perceive various faces learned all through our life expectancy and recognize faces initially even following quite a while of partition. Computational models made for face recognition are very crucial in this modern era as can contribute not only with accurate data and information but also it provides viable applications regarding the same. Computers that detect and recognize faces could be applied to a wide assortment of assignments including criminal recognizable proof, security framework, picture and film handling, character check, labeling purposes and human-computer interaction. The biggest difficulty while building a computational model for face recognition as well as emotion recognition is that the human faces are complex and has different dimensions due to which a meaningful structure is created. The model has to identify the important features in the face and at the same time filter out the errors or unnecessary data from the image. Face identification is currently used in many public places like airport, stations and also particularly on famous sites which facilitate pictures like Picassa, Photobucket, Facebook and Instagram.

The automatically tagging feature adds another aspect to dividing pictures between individuals who are in the image and furthermore gives the plan to others concerning who the individual is in the picture. In our undertaking, we will concentrate on different calculations and strategies with respect to feeling identification and facial acknowledgment and carrying out the most effective and precise calculation in them. Our point is to foster a strategy for emotion detection utilizing face recognition that is fast, robust, reasonably simple and accurate with a relatively simple and easy to understand algorithms and techniques.

2. LITERATURE SURVEY

[1] Emotion recognition is a complex task by nature. As human uses more than one modality to express their emotion, incorporating more than one modality to recognition is not only more natural but also improves the performance of model as one modality acts as complementary to another. In this paper Deep feature fusion base method is used to improve the accuracy of the Emotion recognition to 71%. For multi-modal emotion recognition, for using different modalities as complementary information, feature level fusion is best suitable. By providing raw data as input in audio and visual format, this can be helpful in increasing the recognition rate. However, this process requires more computational resources, data.

[2] The methodology examined is a variation on current ways to deal with eigen image investigation. Contrasted with customary methodologies which use object calculation just (shape invariants), the execution depicted utilizations the eigenspace dictated by processing the eigenvalues and eigenvectors of the image set. The image set is acquired by fluctuating posture while keeping a consistent degree of light in

space, and the eigenspace is processed for each object of interest. For an obscure info image, the recognition calculation extends this image to each eigenspace and the item is perceived utilizing space dividing techniques which decide the article and the situation in space. A few trial results have been gotten to show the heartiness of this technique when applied to the mechanical container picking task.2

[3] In the order cycle, six activity units (AU), determined by the Kinect gadget, were utilized as elements. We utilized closest neighbor classifier (3- NN) and two-layer neural organization classifier (MLP) with 7 neurons in the secret layer. We tried two different ways to perceive feelings: a) subject-subordinate - for every client independently and b) subject-autonomous - for all clients together. In In the two above mentioned scenarios, for 3-NN classifier (Neural Network with 3 prominent learning layers), data were haphazardly segmented into the training part (70%) and the validation part (30%) and for the MLP into three sections: instructing (70%), testing (15%) and approval (15%).

[4] In one of the most notable works in feeling acknowledgment by Paul Ekman, happy, sad, anger, surprise, fear, and disgust were recognized as the six head feelings (other than neutral). Ekman later created FACS utilizing this idea, accordingly setting the norm for work on emotion acknowledgment from that point onward. Neutral was like included later on in most human acknowledgment datasets, bringing about seven essential feelings. Prior works on emotion recognition depended on the customary two venture AI approach, where in the initial step, a few elements are extricated from the pictures and, in the subsequent advance, a classifier (like SVM, neural organization, or arbitrary timberland) is utilized to distinguish the feelings. A portion of the well-known hand-made elements utilized for look acknowledgment incorporate the histogram of situated inclinations (HOG), neighborhood double examples (LBP) , Gabor wavelets, and Haar highlights. A classifier then, at that point, relegates the best emotion to the picture.

3. WORKFLOW

The system will make use of different algorithms in each step. Initially, the system will capture the face of the user in real time using webcam.

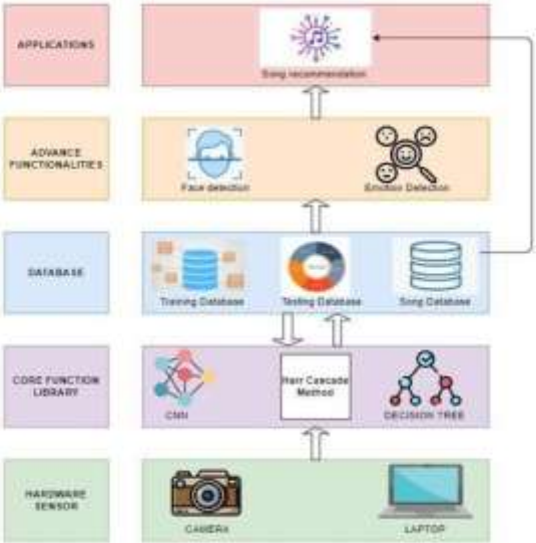

Figure 1. Architecture Diagram

This real time image will be used as an input by Haarcascade_frontalface_default.XML. This file utilizes Haar Cascade Algorithm for face detection. This XML file is available on OpenCV GitHub Repository. After this, the emotion will be classified using CNN (Trained using FER 2013 dataset) and Tensor flow. After emotion is classification, the relevant information is passed on to JavaScript program where the system will fetch the song according to the emotion.

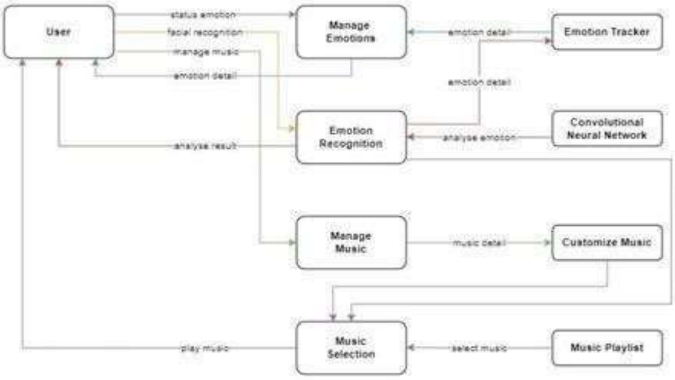

Figure 2. Dataflow Diagram

4. METHODOLOGY

The system will use the Haarcascade_frontalface_default.XML for face detection followed by the emotion detection using CNN model. The emotion detected in the earlier step will then be forwarded to the song playlistalgorithm written in JavaScript and the song will be played accordingly. The application is hosted locally using Django.

We will be using Haar Cascades for face detection due to its low complexity and high performance. Haar Cascade method is one of the most prominent object detection algorithm used for feature extraction in the environment and in our case to detect and identify faces from a real time video or an image. This algorithm uses edge or line detection features proposed by Alfred Haar in 1909 which are also known as Haar features. Utilizing this Haar Features Paul Viola and Michael Jones developed their method which was published in “Rapid Object Detection using a Boosted Cascade of Simple Features” in 2001[9]. Therefore it is also known as Viola Jones Algorithm. This algorithm is divided into 4 stages.

4.1. Harr Cascade Selection

Haar Features are the relevant features for face detection. It was proposed by Alfred Haar in 1909. These features make it easy to find out the edges or the lines in the image, or to pick areas where there is a sudden change in the intensities of the pixels.

For extracting Haar Feature, the algorithm has to perform some calculations which is only possible if the image is gray-scaled. The input from the webcam is a coloured image. Because of this, we require gray-scaled image.

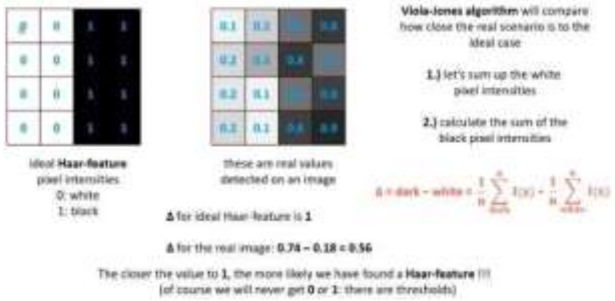
Viola-Jones algorithm will compare how close the real scenario is to the ideal case
1.) let's sum up the white pixel intensities
2.) calculate the sum of the black pixel intensities
ideal Haar-feature pixel intensities
0: white
1: black
these are real values detected on an image
Δ = dark – white = $\frac{1}{n}\sum I(x) - \frac{1}{n}\sum I(x)$
Δ for ideal Haar-feature is 1
Δ for the real image: 0.74 – 0.18 = 0.56
The closer the value to 1, the more likely we have found a Haar-feature !!!
(of course we will never get 0 or 1: there are thresholds)

Figure 3. Haar Feature Calculation

4.2. Creating Integral Images

Now, the haar features traversal on an image would involve a lot of mathematical calculations. This would be a hectic operation even for a highperformance machine. To tackle this, Paul Viola and Michael Jones introduced another concept known as The Integral Image to perform the same calculations more efficiently.

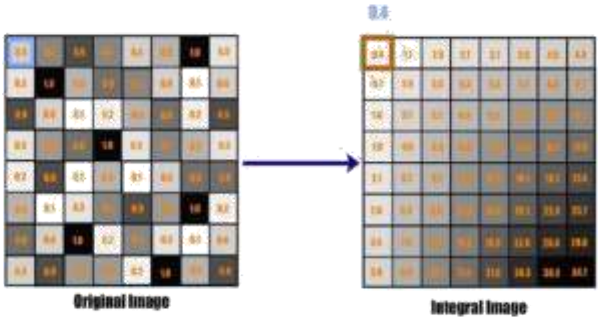

Figure 4. Integral Image Calculation

4.3. AdaBoost Training

Boosting is a method which combines different weak learners and by doing this creates a strong learner. A weak learner is an machine learning algorithm which has accuracy slightly better than random guessing, andstrong learner is the algorithm which has higher accuracy than random guessing. The motive behind boosting methods is to train the weak algorithm iteratively & in each iteration correcting the errors.

4.4. Attentional Cascading

The main motive of attentional cascade is that, there is no need to search each haar feature on each and every window. When the particular haar feature is absent on a certain window, then we can surely say that the facial feature which is to be searched is not present there. Due to this computational time is saved as we can move on to next window.

For emotion classification, we are using the CNN model which is trained on FER 2013 dataset available on Kaggle with the help of Tensorflow library in python.

CNN is a machine learning technique which is developed for processing different types of structured array of data such as images. It is designed to detect patterns and is specialized in pattern detection such as squared,circles and in our case eyes and faces.

Tensorflow is an open-source library in python which provides APIs like keras for model building of CNN. Keras provide easily utilizable APIs due to which number of actions required by user is reduced and this APIs also displays error messages properly.
We used various sub-libraries in keras like:

- *Optimizers*

Whenever a neural network finish processing a batch through the CNN model and generates prediction results, the optimizer calculates the difference between the true value and predicted value and then decide how to use the difference between them. After that it adjust the weights on the node so that the network steps towards the required solution. For this work we are using the Adam Optimizers as it consumes very little memory and is very efficient to use. It also handles a large amount of data which is suitable for our job.

- *Model API*

Model API is the actual CNN model used to add layer, access it(I/Ofunctions) and then sequentialize it.

- *Layer API*

For making shape of the input we are using Layer API.

Conv2D: To produce a tensor of outputs creation of a convolutional kernel is required which done by 2D Convolutional layer.

Dense: Dense is the initial layer which accepts number of neurons orunits in the form of input.

Dropout: To improve the models output we must reduce the overfittingfor which Dropout technique is used. It reduces the error or underperforming models by dropping neurons that don't produce thedesired or near desired results.

Flatten layer: Flatten layer is used to reshape the tensor in such a way that the new tensor has a shape that is equal to number of elements contained in the original tensor.

Input layer: Input layer is basically a tensor and not a layer. It is first hidden layer of the network and also referred to as the starting tensor. This tensor must have the same shape as your training data.

Maxpool 2D: Maxpool 2D is used for operation on 2D spatial Data. The function of Maxpool 2D is to sample down the input along its spatial dimensions (height and weight). This is achieved by taking maximum value over an input window. This process is down for each channel of the input.

Batch Normalization: Batch Normalization is a processing tool which is used to normalize or filter the previous activation layers' output. This technique subtracts the batch mean and then divides it with the batch standard deviation to give the relevent output in the form of feedback to reduce redundancy in the model. For this project, we used the Relu Activation function in Batch Normalization in keras API for the application of rectified linear unit activation function.

- *Image Data Generators*

Image Data Generator class is used for implementing image augmentation. It can generate augmented images dynamically during the training of the model making the overall model more robust and accurate.

- *Callbacks*

Keras API has various callbacks to handle the working of the system.
They are:

Model Checkpoint: Model Checkpoint is a keras callback to save model weights or entire model at a specific frequency. Also, model checkpoint is used whenever a quantity is optimum when compared to last epoch or batch of the network model.

Early stopping: Early stopping is called to stop the training of the model in between or in process. This function is very helpful when your models get overfitted.

ReduceLRonPlateau: ReduceLRonPlateau (Learning Rate) is a essential callback function which reduces the learning rate of the model by some factor (this factor is given by us) whenever the learning stagnates to avoid over learning the model. It is believed that sometimes our model will benefitfrom lowering the learning rate.

Now, the emotion that has been classified will be stored in a variable and then this variable will be passed to a function which will fetch the music/song from the database according to the emotion. These songs will be stored in a custom-made database along with their corresponding titles and song album images which are to be displayed. Also, there will be buttons to navigate between the songs and a detect button to get another emotion.

To support all this, we have created a Django application to bridge between emotion detection system and song recommendation system. Also, Django will provide the local host for user interface where all the results will be displayed.

5. RESULTS AND DISCUSSIONS

As the facial features vary person to person it is difficult to detect human emotions accurately. Still our system has a validation accuracy up to 58-60%, in the initial 15 epochs. The final overall accuracy of the application is increased from 5860% to 72-75%. Compared to other similar papers which have accuracies ranging from 7088% [10], we stand at moderate level of accuracy but with better user interface. The accuracy can be improved if the user uses good quality of camera with adequate lighting. The accuracy of 4 the system mentioned in this paper [11] is about 75% as their system is android based. By following their footsteps and deploying our system on android we can increase the accuracy of the project.

6. LIMITATIONS

The main limitation of the project is that it is hosted on a local system, because of which other people cannot access it using their devices. And also, the accuracy depends on the camera of the device, if the camera quality is good the accuracy will increase. As the songs are stored on local system the overall storage requirement increases.

7. CONCLUSION

Emotion detectors will play a critical part later on and accordingly give an outline of existing methodologies and to comprehend their tradeoffs. This task was embraced to comprehend the face recognition method, analyze the algorithms essential for face detection and analyze the different Machine learning models. In this project, the face recognition and detection algorithms were entirely examined. We additionally concentrated on the Convolutional neural organizations, deep learning and normalization of image. These techniques depend on various strategies for feature extraction, preprocessing and classification along with comparative study. This project was attempted to understand the methodologies of image preprocessing, analyze the features essential for Emotion Recognition. Image processing contributes on developing a computer framework that can perform Image processing Emotion detection and classification techniques are also studied and results will be recorded.

8. FUTURE SCOPE

Further comparative analysis can be made by incorporating different dataset. More clear insights can be made by testing different dataset on proposed multi-model with different AI or machine learning techniques along with using higher pixel cameras and better hardware. Also, on availability of larger dataset, recognition rate can increase without overfitting by adding more layers (CNN blocks) on individual base models

9. ACKNOWLEDGMENTS

I would like to express my sincere thanks to Pune Institute of Computer Technology for giving me the opportunity to write this paper and work on this project. I would like to extend my gratitude to the Department of Information Technology for their continuous support and a very special thanks to my teachers Prof. Manish Khodaskar, Prof. Abhinay G. Dhamankar and Prof. Vinit Tribhuvan, for their constant support and valuable inputs.

10. REFERENCES

[1] Emotion Recognition from Video using feature level fusion by Binod Adhikari, Basanta Joshi in summer 2020.

[2] “Object recognition using eigenvectors” by ‘Ovidiu Ghita, Paul F. Whelan’ 2018, Published by ‘Vision System Laboratory, School of Electronic Engineering, Dublin City University, Glasnevin, Dublin 9, Ireland’.

[3] Emotion recognition using facial expressions by Paweł Tarnowski, Marcin Kołodziej, Andrzej Majkowski, Remigiusz J. Rak, Published in International

[4] Conference on Computational Science, ICCS 2017, 12-14 June 2017, Zurich, Switzerland.

[5] Deep-Emotion: Facial Expression Recognition Using Attentional Convolutional Network by Shervin Minaee, Amirali Abdolrashidi published in 4 Feb 2019.

[6] T. Zhang, "Facial Expression Recognition Based on Deep Learning: A Survey - International Conference on Intelligent and Interactive Systems and Applications ", 2018.

[7] Exploiting multi-CNN feature in CNN-RNN based Dimensional Emotion Recognition on the OMG in-the-wild Dataset by IEEE 10 April 2020.

[8] Racialized emotion recognition accuracy and anger bias of children’s faces. (2020).

[9] Fathallah, L. Abdi and A. Douik, "Facial Expression Recognition via Deep Learning - IEEE/ACS 14th International Conference on Computer Systems and Applications (AICCSA) ", 2017.

[10] “Rapid Object Detection using a Boosted Cascade of Simple Features” by Paul Viola and Michael Jones, 2001,

[11] Music Recommendation System Using Emotion Recognition, IRJET 2021.

[12] Mood Based Music Recommendation System, IJERT June-2021.