ABSTRACT

Music has been a part of life for most of us and some prefer spending hours of time for creating their playlist. Here we propose the help of machine you have with you for almost all day, ‘Smartphones’. There are a lot of automated music recommendation available on a payable subscriptions. We have driven music recommendation to the next level, generating playlist which suits with your mood of listening music. Utilize available resource is the main aspect of this project. All you have to do is to look at your smartphone and listen music. Rest is taken care with the automated tasks. Using front camera of your smartphones, facial emotion is recognized periodically and music playlist is requested using API then songs are played through the music player. This is done so far, this is done so far, the app will also keep track of the songs played, and thus enhancing the recommendations (generate local playlist). So finally after some days of usage, this application will be your best companion of music system.
INTRODUCTION

In this paper, we use machine learning to provide automated playlist with user’s emotion as an attribute. This project is implemented as an Android Application, where the front camera of the smartphone is used to detect the user’s emotion and using it as a parameter in music request to popular services. In addition to it, the application will also keep track of the songs played, so that the future recommendations are improved by this kind of learning is the innovative of the project where everything is automated, and all the user need to do is to enjoy the music. When the user starts enjoying the music the detector will stop to use power.

Human feeling acknowledgment assumes a vital part in the relational relationship. The programmed acknowledgment of feelings has been a dynamic research point from early periods. Accordingly, there are a few advances made in this field. Feelings are reflected from discourse, hand and signals of the body through outward appearances. Subsequently extricating and comprehension of feeling has a high significance of the collaboration amongst human and machine correspondence. This paper depicts the advances made in this field and the different methodologies utilized for acknowledgment of feelings.

The fundamental target of this paper is to propose ongoing usage of feeling acknowledgment framework. The exploration on human PC association (HCI) mode is chiefly to enhance the nature of the collaboration amongst client and PC. Because of contrasts in numerous perspectives, for example, physiology experience and capacity, most current intuitive frameworks can’t address the issues of elderly clients. In view of existing speculations and models the paper joins the examination consequences of inspecting information, offers
contemplations to the attributes of the exceptional gathering elderly individuals, checks Fitt's law, acquires the collaboration mode and outlines rule that can meet its necessities, in this manner empowering all the more elderly clients to accomplish more prominent productivity and fulfillment during the time spent utilizing machine.

**LITERATURE SURVEY**

Taking customary journal keeping as our beginning stage, we have composed and manufactured an advanced journal, named Affective Diary, with which clients can jot their notes, yet that likewise takes into consideration substantial memorabilia to be recorded from body sensors and portable media to be gathered from clients' cell phones.

The framework persistently predicts a client's valence, excitement and draw in ment, and relates this with data on occasions, correspondences and information connections. We assess the interface through a client think about comprising of six clients and more than 240 hours of information, and exhibit the utility of such a reflection apparatus.

Outward appearances, heart rate, breath rate and heart rate inconstancy can advise us about the enthusiastic valence, excitement and commitment of a man. In this proposition I show how consequently identified naturalistic and unconstrained facial reactions and physiological reactions can be utilized to anticipate the adequacy of stories.

We think about the topic of capabilities for hearty visual question acknowledgment, receiving straight SVM based human recognition as an experiment. Subsequent to inspecting existing edge and angle based descriptors, we demonstrate tentatively that frameworks of histograms of situated inclination (HOG) descriptors essentially beat existing capabilities for human identification.
The course can be seen as a protest particular focal point of consideration component which not at all like past methodologies gives measurable ensures that disposed of districts are probably not going to contain the question of intrigue. In the area of face location the framework yields identification rates practically identical to the best past frameworks.

The most recent logical discoveries show that feelings assume a fundamental part in basic leadership, observation, learning, and that's only the tip of the iceberg—that is, they impact the very components of reasonable reasoning. Too much, as well as too little feeling can hinder basic leadership. As per Rosalind Picard, on the off chance that we need PCs to be really astute and to cooperate normally with us, we should enable PCs to perceive, see, even to have and express feelings.

A standout amongst the most vital impacts of feeling lies in its capacity to catch consideration. Feelings have a method for being totally engrossing. Practically, they direct and concentrate on those items and circumstances that have been evaluated as vital to our necessities and objectives so we can manage them suitably. Feeling applicable musings at that point have a tendency to command cognizant preparing—the more imperative the circumstance, the higher the excitement, and the more powerful the core interest.

We depict a dynamic learning approach that helped us proficiently get and hand-mark a huge number of non-nonpartisan unconstrained and common articulations from a huge number of various people. With the expanded quantities of preparing tests an exemplary RBF SVM classifier, generally utilized as a part of outward appearance acknowledgment, begins to end up computationally constraining for preparing and continuous execution. We propose joining two strategies: 1) savvy
determination of a subset of the preparation information and 2) the Nystrom bit guess strategy to prepare a classifier that performs at fast (300fps). We think about execution (precision and classification time) as for the extent of the preparation dataset and the SVM bit, utilizing either a RBF portion, a straight bit or the Nystrom estimate technique.

**PROPOSED METHODOLOGY**

In the proposed system Artificial Intelligence is used for automatic detection of emotion. By using deep learning and neural networks the emotion is detected in an accurate manner. CNN2 is used in the proposed system in order to achieve cost efficiency.

The methodologies are:

- CNN2 is used instead of VGG’16 since VGG’16 uses 3400 flip-flops whereas CNN2 uses 10 times lesser than VGG’16 with same efficiency.

- Face detector shut downs automatically after getting detected and battery efficiency is achieved.

**Architecture**

*There are five modules in this project:*

- Splash Screen Demo
- Dynamic Permissions
- Emotion API
- Music Recommendation
- YouTube API

**SPLASH SCREEN DEMO**

The first time user opens the application, user is treated with a beautiful
introduction about the setup and some basic instructions to use the application. This type of providing introduction to user is called as Splash Screen Introduction, which will run only once (First time user opens the application).

**DYNAMIC PERMISSION**

After Android SDK version 21 introduced, the operating system restricts (Dynamic Permissions) the usage accessing sensitive data from the user’s mobile device. As an application developer, it is a best practice to let the user know why we need particular permission by explaining what we will be doing with those permissions. The Splash screen tutorial we help the user to get to know about the permission and grant them to proceed further.

**EMOTION API**

Emotion API is the important module of the project. Here the emotion detection and processing the camera frames are written in native language (C), which was written with machine learning logic using Neural Networks. JNI is used to access the native language from java (primary for application).

Emotion API consist of following methods:

- *Camera Detector*, initialise camera to send frames to application for processing
- *Detector Listener*, is the event listener which is used to maintain proper life cycle of the Emotion API.
- *Set Detection*, set the required parameters to detect the emotion.
- *Get Detection*, get the emotion detected from the camera frames which is processed.

**MUSIC RECOMMENDATION**

Now, we got the emotion of the user from the Emotion API, the next process is to recommend the music which
will be matched with the user’s current emotion. This can be processed in several ways as many famous music vendors provide API to access their music server using their API. For the project sample purpose, we use YouTube API V3 to request a set videos in the YouTube server. This kind of practice doesn’t affect the view count and the advertisement concept of YouTube as everything is managed by the API provide.

**YOUTUBE API**

The YouTube Android Realer API empowers to join video playback usefulness into Android applications. The API characterizes strategies for stacking and playing YouTube recordings (and playlists) and for tweaking and controlling the video playback encounter. Utilizing the API, we can load or prompt recordings into a player see inserted in the application's UI. We would then be able to control playback automatically. For instance, we can play, interruption, or try to a particular point in the at present stacked video. We can likewise enlist occasion audience members to get call-backs for specific occasions, for example, the player stacking a video or the player state evolving. At last, the API has partner usefulness to help introduction changes and advances to full screen playback.

**ALGORITHM**

**Viola–Jones object detection framework**

The Viola-Jones algorithm is a widely used mechanism for object detection. The main property of this algorithm is that training is slow, but detection is fast. This algorithm uses Haar basis feature filters, so it does not use multiplications.

The efficiency of the Viola-Jones algorithm can be significantly increased by first generating the integral image.

$$\mathcal{L}(y,x) = \sum_{p=0}^{y} \sum_{q=0}^{x} Y(p,q)$$
The integral image allows integrals for the Haar extractors to be calculated by adding only four numbers. For example, the image integral of area ABCD is calculated as

$$II(y_A, x_A) - II(y_B, x_B) - II(y_C, x_C) + II(y_D, x_D).$$

Detection happens inside a detection window. A minimum and maximum window size is chosen, and for each size a sliding step size is chosen. Then the detection window is moved across the image as follows:

- Set the minimum window size, and sliding step corresponding to that size.
- For the chosen window size, slide the window vertically and horizontally with the same step. At each step, a set of $N$ face recognition filters is applied. If one filter gives a positive answer, the face is detected in the current window.

- If the size of the window is the maximum size stop the procedure. Otherwise increase the size of the window and corresponding sliding step to the next chosen size and go to the step 2.
- Each face recognition filter (from the set of $N$ filters) contains a set of cascade-connected classifiers. Each classifier looks at a rectangular subset of the detection window and determines if it looks like a face. If it does, the next classifier is applied. If all classifiers give a positive answer, the filter gives a positive answer and the face is recognized. Otherwise the next filter in the set of $N$ filters is run.

Each classifier is composed of Haar feature extractors (weak classifiers). Each Haar feature is the weighted sum of 2-D integrals of small rectangular areas attached to each other. The weights may take values ±1. Shows examples of Haar
highlights in respect to the encasing location window. Grey areas have a positive weight and white areas have a negative weight. Haar feature extractors are scaled with respect to the detection window size.

The classifier decision is defined as:

\[ C_m = \begin{cases} 1, & \sum_{i=0}^{l_m-1} F_{m,i} > \theta_m \\ 0, & \text{Otherwise} \end{cases} \]

\[ F_{m,i} = \begin{cases} \alpha_{m,i}, & \text{if } f_{m,i} > t_{m,i} \\ \beta_{m,i}, & \text{Otherwise} \end{cases} \]

\( f_{m,i} \) is the weighted sum of the 2-D integrals. \( \alpha_{m,i} \) and \( \beta_{m,i} \) are constant values associated with the \( i \)-th feature extractor. \( \theta_m \) is the decision threshold for the \( m \)-th classifier.

Fig 4.11.1 Object detection Viola-Jones filter

The course engineering is extremely proficient on the grounds that the classifiers with the least highlights are set toward the start of the course, limiting the aggregate required calculation. The most popular algorithm for features training is AdaBoost.

Results And Analysis

The usage of CNN2 improves to efficiency of the emotion detection accuracy. The VGG’16 is 0.74% higher accuracy when compared to CNN2, but the number of MFLOPs required for CNN2 is 3.57 and VGG’16 is 31300. Which enables
CNN2 to be more efficient by detection for mobile devices.

CONCLUSION

In this project, we have introduced automation and machine learning concepts to enhance the music listening experience, where the resource utilization is the primary novelty of the project. As the power and advantages of AI powered applications are trending, our project will be a state of the art of trending technologies utilizations. Our music application will take you to the next level of listening experience which adapts to your mind and seems to be a companion who has the same mood just as you.

References


