International Journal of Early Childhood Special Education (INT-JECSE)
DOI: 10.9756/INT-JECSE/V13I2.211188 ISSN: 1308-5581 Vol 13, Issue 02 2021

EMOTION BASED MUSIC PLAYER USING DEEP LEARNING AND AI

[1]Dr.G.Chitra Ganapathi, [2] P Ponni ,[3] Saranya G
[1]Professor, [2] Assistant Professor, [3]Assistant Professor
Department of Computer Science and Engineering
CMS College of Engineering and Technology,
Coimbatore, Tamil Nadu, India

Abstract: The human face is an important component of an individual's body, and it is particularly useful in determining a person's mood. The essential input can now be extracted straight from the human face using a camera. This data can then be used in a variety of ways. One of the applications of this input could be to extract data in order to determine a person's mood. This information can then be used to generate a list of songs that fit the "mood" produced from the previous input. This reduces the time-consuming and tiresome effort of manually categorising or dividing music into distinct lists and assists in the creation of a suitable playlist depending on a person's emotional characteristics. Several algorithms have been created. Emotions are expressed through words and facial expressions. Written text can also be used to express one's emotions. This paper focuses on what methodologies are available for detecting human emotions in order to develop an emotion-based music player, what approaches are used by existing music players to detect emotions, what approach our music player takes to detect human emotions, and how it is better to use our system for emotion detection. It also gives a quick overview of how our systems work, how playlists are created, and how emotions are classified. The application is thus designed to manage user-accessible content, analyse image properties, and detect the user's mood based on mp3 file properties so that they can be added to the playlist.

INTRODUCTION

The human face is an important part of an individual's body and it especially plays an important role in knowing an individual's mood. Extracting the required input from the human face can now be done directly using a camera. This input can then be used in many ways. One of the applications of this input can be for extracting the information to deduce the mood of an individual. This data can then be used to get a list of songs that comply with the „mood" derived from the input provided earlier. This eliminates the time-consuming and tedious task of manually Segregating or grouping songs into different lists and helps in generating an appropriate playlist based on an individual's emotional features. Various algorithms have been developed and proposed for automating the playlist generation process. Facial Expression Based Music Player aims at scanning and interpreting the data and accordingly creating a playlist based the parameters provided. The scanning and interpreting includes audio feature extraction and classification to get a list of songs belonging to a similar genre or to get a list of similar sounding songs. Human emotions are meant for mutual understanding and sharing feelings and intentions. The emotions are manifested in verbal and facial expressions. One can also express his emotions through written text. This paper mainly focuses on what are the methodologies available for detecting human emotions for developing emotion based music player, which are the approaches used by available music players to detect emotions, which approach our music player follows to detect human emotions and how it is better to use our system for emotion detection. It also gives brief idea about our systems working, playlist generation and emotion classification. The application is thus developed in such a way that it can manage content accessed by user, analyse the image properties and determine the mood of the user based on mp3 file properties so that they can be added into appropriate play lists according to the emotion.

Recognition of facial expressions is used to identify the basic human emotions. Facial expressions give important rules about emotions. Computer systems based on affective interaction could play an important role in the next generation of computer vision systems. Face emotion can be used in areas of security, entertainment and human machine interface. A human can express his/her emotion through lip and eye. Generally people have a large number of songs in their database or playlists. Thus to avoid trouble of selecting a song, most people will just randomly select a song from their playlist and some of the songs may not be appropriate for the current mood of the user and it may disappoint the user. As a result, some of the songs are not matching to the user's current emotion. Moreover, there is no commonly used application which is able to play songs based on the current emotions of the user. Music plays a very important role in enhancing an individual's life as it is an important medium of entertainment for music lovers and listeners and sometimes even imparts a therapeutic approach. In today's world, with ever increasing advancements in the field of multimedia and technology, various music players have been developed with features like fast forward, reverse, variable playback Although these features satisfy the users basic requirements, yet the user has to face the task of manually browsing through the playlist of songs and select songs based on his current mood and behaviour.'

The success of service robotics decisively depends on a smooth robot to user interaction. Thus, a robot should be able to extract information just from the face of its user, e.g. identify the emotional state or deduce gender. Interpreting correctly any of these elements using machine learning techniques has proven to be complicated due the high variability of the samples within each task. This leads to models with millions of parameters trained under thousands of samples. Furthermore, the human accuracy for classifying an image of a face in one of 7 different emotions is 65% ± 5% [4]. One can observe the difficulty of this task by
trying to manually classify the FER-2013 dataset images in Figure 1 within the following classes {"angry", "disgust", "fear", "happy", "sad", "surprise", "neutral"}.
In spite of these difficulties, robot platforms oriented to attend and solve household tasks require facial expressions systems that are robust and computationally efficient. Moreover, the state-of-the-art methods in image-related tasks such as image classification and object detection are all based on Convolutional Neural Networks (CNNs). These tasks require CNN architectures with millions of parameters; therefore, their deployment in robot platforms and real-time systems becomes unfeasible. In this paper we propose an implement a general CNN building framework for designing real-time CNNs. The implementations have been validated in a real-time facial expression system that provides face-detection, gender classification and that achieves human-level performance when classifying emotions. This system has been deployed in a care-O-bot 3 robot, and has been extended for general robot platforms and the RoboCup@Home competition challenges. Furthermore, CNNs are used as black-boxes and often their learned features remain hidden, making it complicated to establish a balance between their classification accuracy and unnecessary parameters. Therefore, we implemented a real-time visualization of the guided-gradient back-propagation proposed by Spring Enberg in order to validate the features learned by the CNN.

LITERATURE SURVEY

DESIGN AND IMPLEMENTATION OF EMOTION RECOGNITION SYSTEM

Human expression plays a vital role in determining the current state and mood of an individual, it helps in extracting and understanding the emotion that an individual has based on various features of the face such as eyes, cheeks, forehead or even through the curve of the smile. Music is basically an art form that soothes and calms human brain and body. Taking these two aspects and blending them together our project deals with detecting emotion of an individual through facial expression and playing music according to the mood detected that will alleviate the mood or simply calm the individual and can also get quicker song according to the mood, saving time from looking up different songs and parallel developing a portable device that can be used anywhere with the help of raspberry pi providing the functionality of playing music according to the emotion detected. Music plays a very primary role in elevating an individual 's life as it is an important medium of entertainment for music lovers and listeners. In today 's world, with the increasing advancements in the field of multimedia and technology, various music players have been developed with features like fast forward, reverse, variable playback speed, genre classification, streaming playback with multicast streams and including volume modulation.

EMOTION BASED MUSIC PLAYING DEVICE

Human expression plays a vital role in determining the current state and mood of an individual, it helps in extracting and understanding the emotion that an individual has based on various features of the face such as eyes, cheeks, forehead or even through the curve of the smile. Music is basically an art form that soothes and calms human brain and body. Taking these two aspects and blending them together our project deals with detecting emotion of an individual through facial expression and playing music according to the mood detected that will alleviate the mood or simply calm the individual and can also get quicker song according to the mood, saving time from looking up different songs and parallel developing a portable device that can be used anywhere with the help of raspberry pi providing the functionality of playing music according to the emotion detected. Music plays a very primary role in elevating an individual 's life as it is an important medium of entertainment for music lovers and listeners. In today 's world, with the increasing advancements in the field of multimedia and technology, various music players have been developed with features like fast forward, reverse, variable playback speed, genre classification, streaming playback with multicast streams and including volume modulation.

REAL-TIME CONVOLUTIONAL NEURALNETWORKS FOR EMOTION AND GENDER CLASSIFICATION

In this paper we propose an implement a general convolutional neural network (CNN) building framework for designing real-time CNNs. We validate our models by creating a real-time vision system which accomplishes the tasks of face detection, gender classification and emotion classification simultaneously in one blended step using our proposed CNN architecture. After presenting the details of the training procedure setup we proceed to evaluate on standard benchmark sets. We report accuracies of 96% in the IMDB gender dataset and
66% in the FER-2013 emotion dataset. Along with this we also introduced the very recent real- time enabled guided backpropagation visualization technique. Guided back-propagation uncovers the dynamics of the weight changes and evaluates the learned features. We argue that the careful implementation of modern CNN architectures, the use of the current regularization methods and the visualization of previously hidden features are necessary in order to reduce

the gap between slow performances and real-time architectures. Our system has been validated by its deployment on a Care-O-bot 3 robot used during RoboCup@Home competitions. All our code, demos and pretrained architectures have been released under an open-source license in our public repository.

DEEP NEURAL NETWORK FOR HUMAN FACE RECOGNITION

Face recognition (FR), the process of identifying people through facial images, has numerous practical applications in the area of biometrics, information security, access control, law enforcement, smart cards and surveillance system. Convolutional Neural Networks (CovNets), a type of deep networks has been proved to be successful for FR. For real-time systems, some pre-processing steps like sampling needs to be done before using to CovNets. But then also complete images (all the pixel values) are passed as input to CovNets and all the steps (feature selection, feature extraction, training) are performed by the network. This is the reason that implementing CovNets are sometimes complex and time consuming. CovNets are at the nascent stage and the accuracies obtained are very high, so they have a long way to go. The paper proposes a new way of using a deep neural network (another type of deep network) for face recognition. In this approach, instead of providing raw pixel values as input, only the extracted facial features are provided. This lowers the complexity of while providing the accuracy of 97.05% on Yale faces dataset. The conventional face recognition pipeline consists of four stages: face detection, face alignment, face representation (or feature extraction), andclassification. The proposed method extracts facial features from input images and feeds them to deep neural networks for training and classification (softmax layer is used). The architecture of network is very flexible and layers can be added or removed to get best results. In recent times there are numerous libraries, functions and platforms to create and modify a network.

JOINT CASCADE FACE DETECTION AND ALIGNMENT

Face detection is one of the mostly studied problems in vision]. The seminal work of Viola and Jones has established the two foundation principles for practical solutions: 1) boosted cascade structure, 2) simple features. Most (if not all) real-time face detectors in academia and industry nowadays are based on the two principles. Such detectors work well for near-frontal faces under normal conditions but become less effective for faces non-frontal or under more wild conditions (lighting, expression, occlusion), because the simple features like Harr in the cascade training are insufficient to capture the more complex face variations. Many works are on multi-view face detection. They adopt a similar divide and conquer strategy: different face detectors are trained separately under different viewpoints or head poses, which are roughly quantized and estimated simultaneously. Because the viewpoint estimation problem is difficult as well and quantization also introduces inaccuracy, such training is more difficult and resulting detectors are usually slower or not accurate enough.

EXISTING SYSTEM

The following are some of the capabilities accessible in existing music players for computer systems: Songs Playlists can be manually selected. Music squares require the user to manually classify songs according to specific feelings for only four fundamental emotions. Passionate, Calm, Joyful, and Excitement are the four emotions. A user had to carefully search through his playlist and select songs that would ease his mood and emotional experience when using standard music players. Various music players have been developed in today's world, with ever increasing advancements in the field of multimedia and technology, with features such as fast forward, reverse, variable playback speed (seek & time compression), local playback, streaming playback with multicast streams, volume modulation, genre classification, and so on. Despite the fact that these characteristics meet the user's basic needs,

DRAWBACKS

- It requires the user to manually select the songs.
- Randomly played songs may not match to the mood of the user.
- User has to classify the songs into various emotions and then for playing the songs user has to manually select a particular emotion.

PROPOSED SYSTEM

Here we propose an Emotion based music player is a music player which play songs according to the emotion of the user. It aims to provide user preferred music with emotion awareness. Player is based on the idea of automating much of the interaction between the music player and its user. The emotions are recognized using a machine learning method Support. In machine learning, support vector machines are supervised learning models with associated learning algorithms that analyse data used for classification and regression analysis. It finds an optimal boundary between the possible outputs.

The training dataset which we used is Olivetti faces which contain 400 faces and its desired values or parameters. The webcam captures the image of the user. It then extract the facial features of the user from the captured image. The training process involves initializing some random values for say smiling and not smiling of our model, predict the output with those values, then compare it with the model's prediction and then adjust the values so that they match the predictions that were made previously. Evaluation allows the testing of the model against data that has never been seen and used for training and is meant to be representative of how the model might perform when in the real world.

According to the emotion, the music will be played from the predefined directories.

FEATURES

- Users don't want to select song manually.
- No need of playlist.
- Users don't want to classify the songs based on the emotions

EXTRACTING EFFECTIVE FEATURES

In this module, first System will capture the image from webcam or suitable device. Then the input image is first checked for the facial features. In case if the image does not contain human features, then it does not detect it. If the input image contains Human features, then it detects the features. Face is detected from image.

FEATURE-POINT DETECTION

In this module, the feature points are detected automatically. For face detection, first we convert binary format image from RGB format image. For converting binary image, we calculate the average value of RGB for each pixel and if the average value is below than 110, we replace it by black pixel and otherwise we replace it by white pixel. By this method, we get a binary image from RGB image. Then next stage is to find the forehead from the binary image. System will start scan from the middle of the image, after that it will look for continuous white pixels after a continuous black pixel. In this we want to find the maximum width of the white pixel by searching vertical both left and right site. Then, if the new width is smaller half of the previous maximum width, then we break the scan because in that case we will reach to the eyebrow. Then we cut the face from the starting position of the forehead and its height will be

1.5 multiple of its width. In this processed image we will only have eyes, nose and lip. Then we will cut the RGB image according to the binary image. This stage can also be achieved using HaarCascade Technology provided by Open CV. HaarCascade handle this with much more optimized and efficient way. Lower face is extracted from.

MUSIC PLAYER

Music Player is developed in java language. It is a simple Music player which includes the functions of pausing the song, next song, previous song and it provides user the facility of managing the database of songs (addition of new songs, Updating of playlists, Deleting song from playlist etc). Playlists will include songs of different categories like the playlist of songs to be played if mood of user is happy or sad or stressed, etc.

DATA ANALYSIS

A face emotion recognition system comprises of two-step process:

i. Face detection (bounded face) in image.
ii. Emotion detection on the detected bounded face.

A face emotion recognition system comprises of two-step process i.e. face detection (bounded face) in image followed by emotion detection on the detected bounded face. The following two techniques are used for respective mentioned tasks in face recognition system. The success of service robotics decisively depends on a smooth robot to user interaction. Thus, a robot should be able to extract information just from the face of its user, e.g. identify the emotional state or deduce gender. Interpreting correctly any of these elements using machine learning (ML) techniques has proven to be complicated due the high variability of the samples within each task. This leads to models with millions of parameters trained under thousands of samples. Furthermore, the human accuracy for classifying an image of a face in one of 7 different emotions is 65% _ 75% [4]. One can observe the difficulty of this task by trying to manually classify the FER-2013 dataset images. In spite of these difficulties, robot platforms oriented to

attend and solve household tasks require facial expressions systems that are robust and computationally efficient. Moreover, the state-of-the-art methods in image-related tasks such as image classification.

TRAINING DATA SET

The data consists of 48x48 pixel grayscale images of faces. The faces have been automatically registered so that the face is more or less centered and occupies about the same amount of space in each image. The task is to categorize each face based on the emotion shown in the facial expression in to one of seven categories (0=Angry, 1=Disgust, 2=Fear, 3=Happy, 4=Sad, 5=Surprise, 6=Neutral). The training set consists of 35,888 examples.train.csv contains two columns, "emotion" and "pixels". The "emotion" column contains a numeric code ranging from 0 to 6, inclusive, for the emotion that is present in the image. The "pixels" column contains a string surrounded in quotes for each image. The contents of this string a space-separated pixel values in row major order.

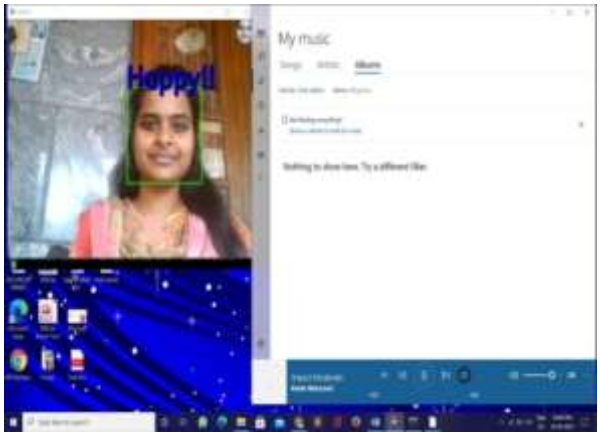
DATA IDENTIFICATION

The usual approach is to split the complete dataset into a training set and a classification set. We use the training set to teach the classifier to recognize the to-be-predicted labels, and use the classification set to estimate the classifier performance. The dataset has been organized and is ready to be recognized, but first we need to actually teach the classifier what certain emotions look like. The usual approach is to split the complete dataset into a training set and a

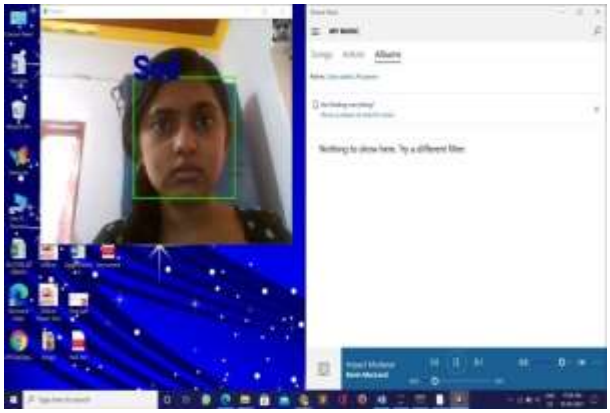
classification set. We use the training set to teach the classifier to recognize the to-be- predicted labels, and use the classification set to estimate the classifier performance.

Estimating the classifier performance on the same set as it has been trained is unfair, because we are not interested in how well the classifier memorizes the training set. Rather, we are interested in how well the classifier generalizes its recognition capability to never-seen- before data. The sizes of both sets depend on what you're trying to classify, the size of the total dataset, the number of features, the number of classification targets (categories). It's a good idea to plot. We'll get into this in another tutorial. For now let's create the training and classification set, we randomly sample and train on 80% of the data and classify the remaining 20%, and repeat the process 10 times. Afterwards we play around with several settings a bit and see what useful results we can get.

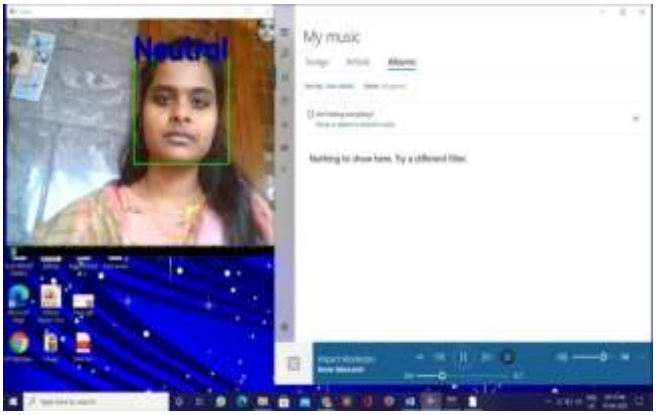
happy Mood

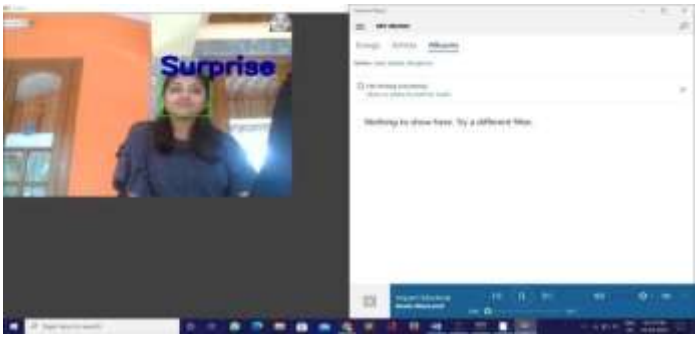


neutral Mood



Fear Mood





CONCLUSION

Music Player has changed in many different ways since it was first introduced. Now-a-days people like to get more out of different applications, so the designing of applications and the thought process behind it has changed. The users prefer more interactive & sophisticated yet simple to use applications. The proposed system (Facial Expression based Music Player) presents a music player capable of generating a playlist from the songs‟ audio features and thereby providing the user with an easy way to get the playlist. The proposed model makes use of the CNN algorithm implemented using java and OpenCV to carry out one phase of its functioning and classifier is used to carry out the audio feature extraction and classification. The references used provide us with vital information about the different techniques and strategies followed to carry out their respective individual systems. Based on the knowledge obtained the above content not only provides an in-depth knowledge of the proposed Software development project system but also tries to incorporate the information from various sources in order to execute the tasks using open-source resources. The various aspects of the project have been presented in the above pages in an adequate manner.

REFERENCES

[1]. Chang, C. Hu, R. Feris, and M. Turk, ―Manifold based analysis of facial expression, Image Vision Comput ,IEEE Trans. Pattern Anal. Mach. Intell. vol. 24, pp. 05–614, June 2006.
[2]. A. habibzad, ninavin, Mir kamalMirnia, A new algorithm to classify face emotions through eye and lip feature by using particle swarm optimization.
[3]. Byeong-jun Han, Seungmin Rho, Roger B. Dannenberg and Eenjun Hwang, ―SMERS: music emotion recognition using support vector regression, 10thISMIR , 2009.
[4]. Alvin I. Goldmana, b.Chandra and SekharSripadab, ―Simulationist models of face-based emotion recognition.
[5]. Carlos A. Cervantes and Kai-Tai Song , ―Embedded Design of an Emotion-Aware Music Player, IEEE International Conference on Systems, Man, and Cybernetics, pp 2528-2533,2013.
[6]. Fatma Guney, Emotion Recognition using Face Images, Bogazici University, Istanbul, Turkey 34342.
[7]. Jia-Jun Wong, Siu-Yeung Cho, Facial emotion recognition by adaptive processing of tree structures.
[8]. K.Hevener,The affective Character of the major and minor modes in music, The American Journal of Psychology,Vol 47(1) pp 103-118,1935.
[9]. Samuel Strupp, Norbert Schmitz, and KarstenBerns, ―Visual-Based Emotion Detection for Natural Man-Machine Interaction.
[10]. Michael lyon and Shigeru Akamatsu, ―Coding Facial expression with Gabor wavelets.,IEEE conf. on Automatic face and gesture recognition, March 2000
[11]. Kuan-Chieh Huang, Yau-Hwang Kuo, Mong-Fong Horng ,Emotion Recognition by a novel triangular facial feature extraction method.
[12]. S. Dornbush, K. Fisher, K. McKay, A. Prikhodko and Z. Segall Xpod- A Human Activity and Emotion Aware Mobile Music Player, UMBC Ebiquity, November 2005.
[13]. Renuka R. Londhe, Dr. Vrushshen P. Pawar, ―Analysis of Facial Expression and Recognition Based On Statistical Approach, International Journal of Soft Computing and Engineering (IJSCE) Volume-2, May 2012.
[14]. Russell, ―A circumplex model of affect, Journal of Personality and Social Psychology Vol-39(6), pp1161–1178 , 1980.
[15]. Sanghoon Jun, Seungmin Rho, Byeong-jun Han and Eenjun Hwang, ―A fuzzy inference- based music emotion recognition system,VIE,2008.
[16]. Thayer The biopsychology of mood & arousal, Oxford University Press ,1989.
[17]. Z. Zeng ―A survey of affect recognition methods: Audio, visual, and spontaneous expressions,IEEE. Transaction Pattern Analysis, vol 31, January 2009.