



Review on Smart Music Player and Algorithm for Recommending Music Based on Facial Micro-Expressions with Multi Cultural Facial Expression Analysis

Dr.A.Rehash Rushmi Pavitra¹ | Anushree K² | Akshayalakshmi A V R² | Vijayalakshmi K²

¹Assistant Professor, Department of Artificial Intelligence & Data Science, Sri Sairam Engineering College, Chennai.

²Department of Artificial Intelligence & Data Science, Sri Sairam Engineering College, Chennai.

*Corresponding Author Email ID: pavitra.ai@sairam.edu.in

To Cite this Article

Dr.A.Rehash Rushmi Pavitra, Anushree K, Akshayalakshmi A V R and Vijayalakshmi K. Review on Smart Music Player and Algorithm for Recommending Music Based on Facial Micro-Expressions with Multi Cultural Facial Expression Analysis. International Journal for Modern Trends in Science and Technology 2022, 8(04), pp. 221-225. <https://doi.org/10.46501/IJMTST0804041>

Article Info

Received: 12 March 2022; Accepted: 06 April 2022; Published: 12 April 2022.

ABSTRACT

Image processing is an emerging technique which performs various operations on an image which in turn in to get a better image or to extract some useful features from it. In general, it is a type of signal processing in which the input is an image and output is also image together with characteristics or features associated with the image. In addition to this, "songs" meant to a medium of expression that have always been the best choice to analyse and understand human emotions. A mixture of micro-expression recognition technology of convolutional neural network and systematic music recommendation algorithm is developed into a model that recognizes facial micro-expressions with multicultural facial expression details and recommends music according to corresponding mood. To achieve classify facial expressions into five different emotional categories such as happy, sad, angry, neutral and disgust. Further, proposed techniques incorporate "Hand-gestures" as an extra feature. The main goal of this research paper is to disclose the summary of an effective music player and social companion which automatically generate a mood-lifting playlist based on the emotion state of the user together suggesting future research in the area of recommendation system.

KEYWORDS: Facial recognition, Artificial intelligence, Machine learning, CNN, FER, Recommendation system.

1. INTRODUCTION

Music can be a medium of pleasure and satisfaction but there are other psychological benefits as well. Listening to music has been proved results to improve memory processing, increase healing time, improves our fitness mode and so on. Music therapy is a therapeutic approach which uses instinctively mood-lifting characters of music to help patients to improve their own

mental health and overall well-being. The history behind the advantages and applications of music therapy has been researched for recent days. Major consequence from analytical studies have appeared to be that music therapy may be supportive for people with depression, sleep disorders, anxiety and even healed cancer.

In pandemic period, the primary cause of stress arises due to with quarantine-at-home that made many people

reaching for music as an instrument to help them to fight their tough times. In the US, Meta Group called "Quarantine Karaoke" attracted more than half a million people who sang songs for each other on social media. Hence, many used music to battle loneliness.

The presence of vast networks in the past decades, internet became the major root of recover multimedia data such as books, music and videos etc. Considering, the fact that music is an important prospect of their lives and they listen to music as a regular activity. The very important part in music listening is the emotional state of a person who is listening to music which can influence the type of music being selected for listening.

Image processing:

Image processing is the process of working on some tasks in an image either for extracting useful information from the images or to get a refined image. In this process the input is an image and the output may be image or features associated with the image. Image processing involves three steps:

1. Importing the picture by using image detection tools.
2. Scrutinize and make changes the image.
3. The result produced by the image or a report based on analysing the image can be changed.

Image processing methods are of two types. In analogue image processing hard copies like photographs, printouts are used. Whereas in digital processing, processes the digital images by using the digital computer. It has many advantages compared to analogue processing, because it allows larger number of algorithms. And during pre-processing it avoids problems such as noise and distortion. Hence, digital image processing is used in many real time applications.

Applications:

Image processing has many applications in our day-to-day life. In video processing this method is used for noise removal, motion detection, brightness and contrast. In medical sector image processing is seen in for imaging of X ray, UV imaging, Gamma ray imaging etc. The typical functions of image processing are converting the gray scale to color image, zooming, blurring, sharpening, image detection, Image recognition. In pattern recognition this method is used to recognize various patterns. In robotics field image processing is used for hurdle detection to increase the vision of the robot and also useful for line follower

robots. In transport sector it is used for identification of numbers and traffic signs automatically.

Advantages:

The digital image can be made available in any format. Images can be stored in computer memory and can be retrieved easily at any time. By using image processing it can easily change the size of the pixel in the image to desired size and density. Image Processing made the digital image noise free. It allows electronic transmission of images to third parties.

Disadvantages:

The images can be copyrighted. The cost of the system used for this method may be high. If the computer crashes the images cannot be retrieved. The changes in lighting, background pictures can affect the face detection.

Feature Extraction:

Feature extraction is a part of dimensionality reduction process. It is the process of reducing the initial set of raw data to a more manageable group that can be processed. It gives better results compared to applying the machine learning algorithms directly to the raw data. Feature extraction is the crucial step in image processing. In image processing feature extraction is important in optical character recognition. Further, python language is used for feature extraction in distinct resources. Thereby, use of feature extraction is to obtain the important features of a face (eyes, lips, nose) as well hand motions.

Convolution Neural Network (CNN):

CNN refers to convolution neural network here more complex features of the image can be extracted and learnt by it. It consists of a combination of two functions in order to obtain a third function and it is a mathematical operation. In this algorithm, convolution is performed on input data by adding a convolution layer with the help of filters to produce a feature map. After that pooling layer is added this is used to reduce the computation and number of parameter. As a consequence of this it controls the overfitting problem and reduces the training time. Max-pooling is the widely used polling technique. Ultimately, it takes the highest value in each window that reduces the feature map size, however keeping the significant information.

Data mining:

Data mining is the method of extracting and uncovering pattern behind the huge datasets and transforming it into comprehensible structure for future use. Data mining depends on effective data warehousing, computer processing and data collection. Data mining is used inspam email filtering, database marketing, fraud detection, credit risk management, or to get the opinion from the customers. It relies on big data, machine learning and artificial intelligence to predict the outcomes from large data sets. There are four main steps in this process.

1. Setting the objective.
2. Gathering and preparation of the data.
3. Applying the data mining algorithm.
4. Assessing the result.

Here, data mining for lyrics analysis and extract the sentiment and emotions from those words for recommending songs based on lyrics is used accordingly.

Lexicon:

The words used in a language or by a person or group of people.

Lexicon -Sentiment Analysis:

This process works on matching a sentiment word or phrase using a lexicon as the computerscannot understand human language. This method is a crucial one in analysing sentiment in text mining. Each and every word has some weight value and labelled as positive or negative.

Lexicon - Emotion Analysis by 'Paul Ekman':

Paul Ekman, an American psychologist, explained very well in his paper saying that humans' emotions can be classified into six i.e., joy, fear, anger, sadness and disgust. This method allows us to distinguish emotions more accurately than the sentiment analysis such as negative and positive, used in "traditional" text mining.

Data sets and Databases:

A data set is a collection of data. The data set is generallyrepresented in a tabular form. The columns and rows represent a particular variable and values respectively. The database is very important for image

processing. The selecting of the dataset is significant in facial recognition. The following important data sets for training are listed below:

The Japanese female facial expression (JAFFE) is a dataset composed of 213 face images of 10 female Japanese and has a resolution of 256 x256. The images are in gray colour. Taiwanese Facial Expression Image Database (TFEID), this database consists of 7200 stimuli captured from 20 males and 20 females. It shows eight facial expressions. Radboud Face Database (RaFD) comprise of 838 images from 19 female and 38 male participants. Facial Expression Recognition 2013 Dataset (FER2013) contains approximately 30,000 images with size restricted to 48x48 pixel. It is organized with seven motion classes. For efficient training furthermore database will be used in future as well.

2. REVIEW OF LITERATURE

A blend of micro-expression recognition technology of convolutional neural network and automatic music recommendation algorithm is reputable to identify a model that distinguishes facial micro-expressions and recommends music rendering to corresponding mood [1]. The facial micro-expression recognition model established in this published paper uses FER2013 with a recognition rate of 62.1% [6]. The authors produced a moving cross-platform music player, called EMP, which recommends music based on the actual mood of the user. EMP provides smart mood-based music recommendation system, containing 3 modules: emotion, music classification and recommendation modules. EMP diminishes user exertions for producing playlists by efficiently plotting the user's emotion to the correct song class with an overall accuracy of 97.69%, it achieves hopeful results for the four moods studied. In this study, new convolutional neural network architecture, namely, PyFER, is projected to address the FER problem [7]. The experimental results established that the proposed neural network architecture is fast enough to be integrated into present FER applications as it was able to complete the analysis of a given picture for an average of 12.8 milliseconds, which is in the endurable limit to latency for real-time applications and also the accuracy of PyFER was calculated to be as high as 96.3% on a de-facto standard dataset, namely, CK+, and all facial expressions, except for happiness, were

correctly detected by PyFER. They executed an effective method to detect faces and emotions of the people [3]. For detecting the face from the picture, they have used the well-known Viola Jones face detection method and for detecting voice features “Mel” frequency components of the human voice are used. By using KNN classifier algorithm, recovered face and emotion restructuring of the person is identified. Experimental results casted that the efficiency of the proposed face and emotion restructuring system is 94.5 to 97 %. In their prototype, the user is able to listen to current mood playlists available in the system, also listen to the playlists created by other users of the same system, and/or upload new images as the source for creating new playlists [10]. Music can then be associated to the uploaded pictures by all the users of the system. The playlist creation is a communicating process and the playlists are continuously evolving and receiving new add-ons over time. Planned a method for music recommendation using emotion and sentiment analysis based on the lyrics using “text mining”, performed the refining process using “Stopword” and Natural Language Processing, and analysed the lyrics of music based on lexicon, used NRC Emotion Lexicon; dividing lyrical words into sentiment and emotion parts [14]. After selecting a new song using the K-Nearest Neighbour (K-NN) algorithm, the authors proposed a method for recommending music with the most similar lyrics. Established a prototype called ‘EmoPlayer’, an Android application, which helps to seizure user’s image using camera and detects his face, along with the emotions and creates a list of songs which will enhance his/her mood as the songs keep playing behind [11]. Even this player uses Viola Jones algorithm for face detection and uses “Fisherfaces” classifier for emotion classification. This system uses 450 images to train the classifier which further used 80% of the image set and tested for its accuracy using the rest 20%. This work examines music recommendation method based on dance movements, uses 3D human skeleton data for music recommendation [15]. Experimental results showed that the proposed method achieves a high accuracy (91.30%) for music recommendation, which outdoes the music generation method (0% for per music piece accuracy, and 16.67% for per music genre accuracy) with a large margin.[2] Proposed an artificial neural network-based ensemble classifier for

“multicultural” facial expression analysis. The facial images from the Rad Bound database, Taiwanese facial expression image database, and Japanese image database are combined to form a combined multi-culture facial expression dataset. The authors studied and proposed an automatic playlist generation method which analyses self-made playlists, recognizes their structure and development and generates new playlists accordingly [4]. Adopted Recurrent Neural Network (RNN) for the sequence modelling. Furthermore, since the representation model adopted to describe each song is basis and is also connected to the human insight, rewards of Convolutions Neural Network (CNN) is used to learn meaningful audio descriptors. The authors proposed an efficient facial expression recognition algorithm merging appearance and geometric features based on deep neural networks for more precise and effective facial expression recognition [12]. The proposed algorithm produced 91.3% of the accuracy which was improved by 1.5% when compared with other existing methods in the JAFFE dataset.

This study proposes the recognition of micro-expression emotions with estimated subtle motion in the image sequence based on onset-frame, apex-frame, and offset-frame for feature extraction [8]. The method used for feature extraction is the combination of Block Matching Algorithm with Taylor Series Approximation referred to as Subpixel Subtle Motion Estimation (SME). Multiclass Classification process using Multilayer Perceptron (MLP) Backpropagation and Support Vector Machine (SVM) as a comparison. The evaluation results show the best accuracy 85.07%, with the Mean Absolute Error 0.0597 and Root Mean Square Error 0.2443. Implemented a method to collect both unspoken feedback and textual features collected from YouTube to improve the recommendation performance [5]. Experiment results show that the recommendation playlists generated by this system matches both individual’s and group preference. A mobile device has limited processing power; the algorithm in the emotion recognition system should be implemented using less computation. In this paper, the authors proposed emotion recognition with high and satisfied performance for mobile applications [13]. In the proposed system, facial video is captured by a rooted camera of a smart phone. A depth camera-based novel method is proposed here for efficient facial expression

recognition [9]. For each pixel in a depth image, eight local directional strengths are obtained and ranked. Once the rank of all pixels is obtained, eight histograms are processed for the eight surrounding directions. The histograms are then combined to represent features for a depth image of a face. This approach is named “local directional rank histogram pattern”.

3. CONCLUSION

The proposed review focuses with image processing and facial emotion recognition system, human’s emotions can be understood in depth and with its uses and applications, can be used in healthcare sector, especially in music therapy as well as a good social companion for a loner. Recommendation system too plays a role by creating better options or choices by learning the structure behind a person’s sentiments. Also recognize the chance for improvement. It would be fascinating to analyse how the model performs when all seven emotions are taken with equal score of implementations i.e., same accuracy scores into consideration. Even though we could limit the cultural barrier, the language barrier for songs is still not up to our expectation. Extra songs from different regional languages can also be added to make the recommendation system more robust.

Conflict of interest statement

Authors declare that they do not have any conflict of interest.

REFERENCES

[1] Ziyang Yu, Mengda Zhao, Yilin Wu, Peizhuo Liu and Hexu Chen, “Research on Automatic Music Recommendation Algorithm Based on Facial Micro-expression Recognition”, Proceedings of the 39th Chinese Control Conference July 27-29, 2020, Shenyang, China.

[2] Ghulam Ali, Amjad Ali, Farman Ali, Umar Draz, Fiaz Majeed, Sana Yasin, Tariq Ali and Noman Haider “Artificial Neural Network Based Ensemble Approach for Multicultural Facial Expressions Analysis”, Received June 27, 2020, accepted July 7, 2020, date of publication July 17, 2020, date of current version August 3, 2020.

[3] Dolly Rene, Dr. Neeta Tripaathi, “An Efficient Method to Face and Emotion Detection”, 2015 Fifth International Conference on Communication Systems and Network Technologies.

[4] Rosilde Tatiana Irene, Clara Borrelli, Massimiliano Zanoni, Michele Buccoli and Augusto Sarti, “Automatic playlist generation using Convolutional Neural Networks and Recurrent Neural

Networks”, 2019 27th European Signal Processing Conference (EUSIPCO).

[5] Yi-Chen Wang, Pei-Lin Yang, Sok-Ian Sou and Hsun-Ping Hsieh, “The MuTube Dataset for Music Listening History Retrieval and Recommendation System”, 2020 International Computer Symposium (ICS).

[6] Shlok Gilda, Husain Zafar, Chintan Soni and KshitijaWaghurdekar, “Smart Music Player Integrating Facial Emotion Recognition and Music Mood Recommendation”, This full-text paper was peer-reviewed and accepted to be presented at the IEEE WiSPNET 2017 conference.

[7] Abdullah Talha Kabakus, “PyFER: A Facial Expression Recognizer Based on Convolutional Neural Networks”, Received July 21, 2020, accepted July 26, 2020, date of publication July 29, 2020, date of current version August 14, 2020.

[8] Niyalatul Muna, Ulla DelfanaRosiani, EkoMulyantoYuniarno and MauridhiHery Purnomo, “Subpixel Subtle Motion Estimation of Micro-Expressions Multiclass Classification, 2017 IEEE 2nd International Conference on Signal and Image Processing.

[9] MD. Zia Uddin, (Senior Member, IEEE), Weria Khaksar and Jim Torresen, (Senior Member, IEEE), “Facial Expression Recognition Using Salient Features and Convolutional Neural Network, Received September 26, 2017, accepted October 23, 2017, date of publication November 23, 2017, date of current version December 22, 2017.

[10] ArtoLehtiniemi and Jarno Ojala, “Evaluating MoodPic – A Concept for Collaborative Mood Music Playlist Creation, 2013 17th International Conference on Information Visualisation.

[11] Aurobind V. Iyer, K J Somaiya COE, Viral Pasad, Smita R. Sankhe and Karan Prajapati, “Emotion Based Mood Enhancing Music Recommendation”, 2017 2nd IEEE International Conference On Recent Trends in Electronics Information & Communication Technology (RTEICT), May 19-20, 2017, India.

[12] Ji-hae Kim, Byung-Gyu Kim, (Senior Member, IEEE), ParthaPratim Roy, (Member, IEEE), and Da-mi Jeong, “Efficient Facial Expression Recognition Algorithm Based on Hierarchical Deep Neural Network Structure”, Received February 22, 2019, accepted March 20, 2019, date of publication March 25, 2019, date of current version April 11, 2019.

[13] M. Shamim Hossain, (Senior Member, IEEE), and Ghulam Muhammad, (Member, IEEE) An Emotion Recognition System for Mobile Applications, Received January 30, 2017, accepted February 8, 2017, date of publication February 22, 2017, date of current version March 15, 2017.

[14] Jinhyuck Choi, Jin-Hee Song, Yanggon Kim, “An Analysis of Music Lyrics by Measuring the Distance of Emotion and Sentiment”, SNPD 2018, June 27-29, 2018, Busan, Korea.

[15] Wenjuan Gong, (Member, IEEE), and Qingshuang Yu “A Deep Music Recommendation Method Based on Human Motion Analysis”, Received December 29, 2020, accepted January 26, 2021, date of publication February 5, 2021, date of current version February 17, 2021.