EasyChair Preprint

№ 10645

Music Player Using Emotion Detection

Bhargava Kumar Gollamoni, Rohit Babu Grandhi,
Pranay Kumar Gurram, P Subhadra and G Srikanth Reddy

August 1, 2023

MUSIC PLAYER USING EMOTION DETECTION

G. Bhargava Kumar
bhargava.naidu01@gmail.com
Computer Science and Engineering
VBIT, Hyderabad

G. Rohit Babu
grohit0303@gmail.com
Computer Science and Engineering
VBIT, Hyderabad

G. Pranay Kumar
gurrampranay143@gmail.com
Computer Science and Engineering
VBIT, Hyderabad

MRS.P. SUBHADRA
Associate Professor
VBIT, Hyderabad

Mr. G. Srikanth Reddy
Associate Professor
VBIT, Hyderabad

ABSTRACT - Human emotion plays a vital part in recent times. Emotion is grounded on Human passions which can be both expressed or not. Emotion expresses the human's individual state which can be in different forms. Extraction of the emotion states humans individual state of mood. Facial expressions are captured an original capturing device or an inbuilt camera. Then we use Haar Cascade algorithm for the recognition of the point from the captured image. Therefore, the proposed system is grounded on the facial expression captured. The Fisher Face algorithm is used to descry different mortal feelings and music will be played automatically.

Keywords — Facial expression, Emotion recognition, Haar Cascade Algorithm, Fisher Face Algorithm.

I. INTRODUCTION

One's life is significantly impacted by music. It serves as a significant form of entertainment and is frequently used therapeutically. An individual's life includes music in a significant way. It serves as a significant source of amusement and is frequently used as a corrective measure. The introduction of technology and ongoing developments in multimedia have led to the creation and amendment of sophisticated music players with a variety of capabilities, including volume modulation. In spite of the fact that this point effectively addresses the user's desire, the user still needs to search his playlist for music that express his emotions. In a conventional music player, the user is required to independently search through his playlist and choose songs that would enhance his mood and emotional experience. This song selection process is difficult and time-consuming, and the user may have trouble finding a song that fits their needs. The introduction of Music Information Retrieval (MIR) and Audio Emotion Recognition (AER) gave the conventional systems a point that automatically parsed the playlist based on several types of emotions. In the current system, the user chooses the song he wants to hear each time, so we developed a system that uses the user's emotions to generate melodies. Users get bored when they hear the same melody repeatedly, therefore we create a different tune for each time a user logs in. Every time the same human emotion is identified, the system produces the same melody. The user also grows weary of hearing the same song over. Therefore, in our system, we use a system that will generate a new tune as a development of a previously generated tune. Therefore, a different melody will be produced each time the same face expression is detected.

II. RELATED WORKS

[1]S. Deebika, presented a paper on "A Machine Learning Based Music Player by Detecting Emotions." The paper constitutes the implementation of Convolutional neural network for the emotion detection and thereby playing a song accordingly. Segregating the songs and playing them in accordance to one's mood could facilitate the music lover. Although there exist a lot of algorithms designed for it, the computation is not as expected. This paper eradicates such an issue by using CNN. In order to obtain minimal processing, multilayer perceptron is implemented by CNNs. In comparison to various algorithms for image classification, CNNs observed to have little-processing. This implies that the filters used in CNNs are advantageous when compared to traditional algorithm. The visualization of features directly can be less informative. Hence, we use the training procedure of back-propagation to activate the filters for better visualization. The multiple

actions such as capturing, detecting the emotion and classifying the same can all be confined as one step through the use of CNN. The slow performances of the real-time approaches could be enhanced by regularizing the methods and by visualizing the hidden features. Hence the proposed approach could enhance the accuracy and the computation speed.

[2] K. Patel and R. K. Gupta, "Song Playlist Generator System Based on Facial Expression and Song Mood". Because of the hectic pace that people have nowadays, life is incredibly hectic. People are increasingly inclined to listen to music while performing their daily duties which help them relax after a stressful day. As a result, songs become important part of daily lifestyle. Due to the huge demand several music players have entered to the market and try to attempt to deliver the best possible music recommendation for the customer. This paper proposes a Deep Learning based approach for the playlist generation based on human current mood with the help of user's past history of song selection. In this approach we are trying to generate playlist from the emotion of the user to add touch of current situation of user mood and user personal choices of the songs for providing more personalized experience. After introduction of the Convolutional Neural Network object detection, Image classification, Emotion detection tasks reaches great height. In the proposed method, we use convolution neural network (CNN) for emotion detection task and artificial neural network (ANN) for song classification task. Experiment result says that our suggested model achieves 84% accuracy on FER-13 dataset which contain around 14k facial images. For song classification task we have used different song-features which is extracted from Spotify music player. We have achieved 82% accuracy in song classification task. Currently this system is only with Spotify music player. Motivation of this approach is to provide better song recommended playlist based on user current mood.

[3] J. H. Bae, M. Kim and J. S. Lim, "Emotion Detection and Analysis from Facial Image using Distance between Coordinates Feature." Facial expression recognition has long been established as a subject of continuous research in various fields. In this study, feature extraction was conducted by calculating the distance between facial landmarks in an image. The extracted features of the relationship between each landmark and analysis were used to classify five facial expressions. We increased the data and label reliability based on our labelling work with multiple observers. Additionally, faces were recognized from the original data, and landmark coordinates were extracted and used as features. A genetic algorithm was used to select features that were relatively more helpful for classification. We performed facial recognition classification and analysis using the method proposed in this study, which showed the validity and effectiveness of the proposed method.

[4] K. M. Kudiri, A. M. Said and M. Y. Nayan, "Emotion detection through human facial expressions using relative grid-based coefficients." Face-to-face interactions with computers involve the use of facial expressions so that the computer can read human emotions and respond appropriately. In interactions between people or between people and computers, facial expressions serve as a nonverbal communication cue. It is challenging to obtain visual information in noisy environments. High dimensionality is hurting the system, which in turn is affecting the performance of the emotion detection system, according to relative bin sub-image-based studies. These factors led to the development of a novel method leveraging relative grid coefficient feature extraction from visual data. For the classification of emotions, radial basis kernel support vector machines are utilised. According to preliminary findings, relative grid-based characteristics were retrieved with an average accuracy of 89%.

III. SYSTEM MODEL

This is a design grounded on the conception of Computer Vision and Emotion discovery. In this design, we will use computer vision to describe the emotion a person is flaunting. Through the emotion detected, the system shall identify a song playlist according to the mood of the user. It'll give easy access to any song the user wants to hear according to their mood therefore reducing the work. The design is grounded on the technology of computer vision and specifically on the subfield of Emotion discovery. This technology is used in collaboration with the generation of asking music, based on the mood detected.

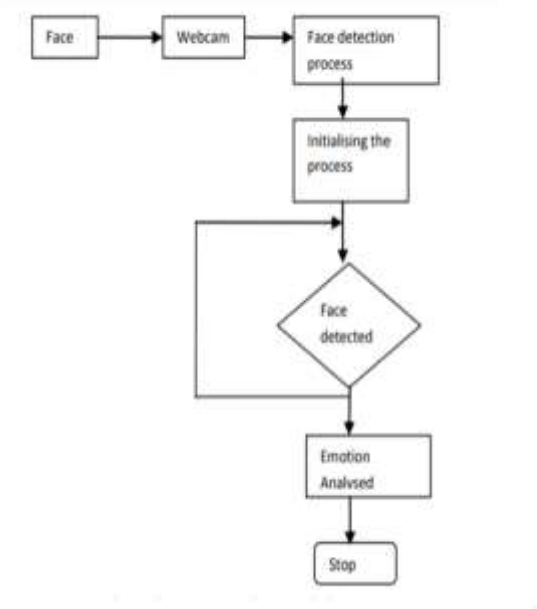

Fig Architecture diagram

IV. PROPOSED MODEL

A. Creating the Dataset

i. Images:

Images are a rich source of information, and using them as a dataset can be extremely valuable in machine learning projects. In this context, the images are typically used as input data, which the machine learning algorithm can use to learn patterns and make predictions or classifications. When working with image datasets, it's important to consider factors such as image resolution, size, and format, as well as how to preprocess the images before feeding them into the machine learning model.

One popular approach to working with image datasets is to use convolutional neural networks (CNNs). These networks are specifically designed for image recognition tasks and can identify patterns in the visual data by analyzing groups of pixels together. The CNNs can be trained using supervised or unsupervised learning techniques, depending on the availability of labeled data.

To use images as a dataset for a machine learning project, it's typically necessary to have a large number of images that are relevant to the problem at hand. For example, if the project is focused on identifying cats and dogs in images, a dataset of thousands of cat and dog photos would be necessary to train the model effectively.

Overall, the use of images as a dataset in machine learning projects is a powerful tool for solving a wide range of problems. With the right techniques and tools, it's possible to extract valuable insights from visual data and use this information to make predictions or classifications with high accuracy.

Our data set consists of several images maintaining information regarding the face of human beings. These images are segregated by us with several emotion labels like angry, sad, happy, neutral etc. It can be used by the model to train itself by the knowledge presented to it.

Fig Sad dataset

Fig Angry dataset

ii. Songs:

Music has the power to evoke emotions, and choosing a playlist of songs according to emotion is important because it can enhance our emotional state and improve our overall well-being. Music is known to affect our brainwaves and can stimulate the release of dopamine, a feel-good neurotransmitter, which can help to boost our mood and reduce stress.

By selecting songs that match our emotional state, we can tap into the therapeutic benefits of music and create a personalized playlist that helps us to manage our emotions. For example, if we're feeling sad or anxious, listening to soothing music can help to calm us down and provide a sense of comfort. Similarly, if we're feeling happy and energetic, upbeat music can help to amplify these positive emotions and improve our motivation.

Moreover, choosing a playlist of songs according to emotion is important for specific occasions. For instance, a workout playlist would require high-energy, motivating songs to boost our energy levels and keep us motivated during exercise. On the other hand, a relaxing playlist before bedtime would require calming and gentle songs to help us unwind and prepare for sleep.

In summary, choosing a playlist of songs according to emotion is important for our emotional well-being and can provide therapeutic benefits to improve our mood and reduce stress. By selecting music that matches our emotional state, we can create a personalized playlist that enhances our emotional experience and helps us to manage our emotions effectively.

To enhance the system our data set consists of various music files (presented in .mp3 format). This data set is categorized based upon the mood depicted by that song. This will be played after detecting the emotion.

B. Capturing the image

The primary process of our project includes the live capturing of the face of user, using the web cam interface. This is achieved by using the OpenCV methods by the python libraries.

OpenCV is a powerful computer vision library that provides a wide range of tools for processing and analysing images and video. One of the key features of OpenCV is its ability to capture live images from various sources, such as webcams and cameras. This functionality can be accessed using the Video Capture class, which allows users to access and process live video streams in real-time. By using OpenCV's live image capture capabilities, developers can create a wide range of applications, from video surveillance systems to augmented reality experiences. Overall, OpenCV's live image capture capabilities make it a valuable tool for anyone working with computer vision and image processing applications.

The captured image is analysed by the model to extract the features over the facial image. By using Haar cascade algorithm we will be extracting the features of the image captured in the live model.

The Haar Cascade algorithm works by first detecting the edges or features of an object in an image. These features are typically rectangular regions of the image that differ in intensity from their surroundings. In the case of face detection, common features include the edges of the eyes, nose, and mouth.

Haar Cascade uses a pre-trained classifier, which is essentially a set of rules that tell the algorithm how to identify faces. The classifier is created by training a machine learning model on a large set of positive and negative examples of faces. During training, the model learns to distinguish between positive examples (images of faces) and negative examples (images without faces).

Once the classifier is trained, Haar Cascade uses a sliding window approach to scan an input image for potential face regions. The sliding window is moved across the image, and at each position, the classifier evaluates whether the window contains a face or not. If the classifier detects a face, the position and size of the face are recorded. It is designed to be computationally efficient, allowing it to process images and video frames in real-time.

C. Training the model

The main process of our model starts by executing the .xml file which will be redirecting the images (i.e., emotions dataset) to the training model and build the model. The images presented will be trained to differentiate the various emotions presented by the user. By using the machine learning algorithms, we can achieve this step and improve our accuracy day by day and model by model.

Fisher face algorithm is a popular technique used in face recognition to extract discriminant features from a given dataset. It uses linear discriminant analysis (LDA) to transform the dataset into a lower-dimensional subspace, where the distances between different classes of data are maximized, while the distances within each class are minimized. This results in a set of features that are optimized for classification and can be used to train a machine learning model.

To train a model using Fisher face algorithm, the first step is to prepare the dataset. This involves collecting images of faces and labelling them according to their respective classes. For example, if we are building a model to recognize faces of different people, we would need to collect images of each person and label them accordingly.

Next, we need to pre-process the images by normalizing them and removing any noise or background information. This can be done using techniques such as histogram equalization, which ensures that the image intensities are spread evenly across the entire range.

Once the dataset is prepared, we can use the Fisher face algorithm to extract discriminant features from the images. This involves calculating the eigenfaces of the dataset, which represent the principal components of the dataset. These eigenfaces are then used to project the original images onto a lower-dimensional subspace, where the distances between different classes are maximized.

The resulting features can then be used to train a machine learning model such as a support vector machine (SVM) or a neural network. The model is trained using a subset of the dataset and tested on the remaining data to evaluate its performance. If the model performs well, it can be used to recognize faces in real-world applications.

Overall, training a model using Fisher face algorithm can be a powerful tool in face recognition applications. By extracting discriminant features from a given dataset, we can create a model that can accurately recognize faces and classify them into different classes. With the right techniques and tools, we can build robust and accurate models that can be used in a wide range of real-world applications.

D. Emotion analysis

Emotion analysis and recognition is a rapidly growing field in computer vision and machine learning. With the availability of large datasets and advanced algorithms, it's becoming increasingly possible to accurately detect and recognize human emotions from live images.

One approach to emotion analysis and recognition is to use a trained dataset of facial expressions. This involves collecting a large number of images of faces expressing different emotions, such as happiness, sadness, anger, and surprise, and using machine learning algorithms to train a model to recognize these emotions. The model is then used to analyse live images in real-time and identify the emotional state of the person in the image.

To achieve this, the live images are first pre-processed to remove any noise or irrelevant information, and then features such as the position and shape of facial landmarks are extracted using algorithms such as facial landmark detection. These features are then fed into the trained machine learning model, which can recognize the emotional state of the person in the image with a high degree of accuracy.

Emotion analysis and recognition using live images has a wide range of potential applications, from marketing and advertising to healthcare and psychology. For example, it can be used to measure the emotional response of users to different products or services, or to help diagnose and treat mental health disorders by analyzing facial expressions and emotional states.

However, there are also challenges associated with emotion analysis and recognition using live images, such as the need for large and diverse datasets to train the machine learning models, and the potential for biases in the data that can affect the accuracy of the results. Additionally, ethical considerations such as privacy and consent need to be taken into account when collecting and analysing live images.

In summary, emotion analysis and recognition using live images and trained datasets is a rapidly advancing field that has the potential to revolutionize a wide range of industries. By accurately detecting and recognizing human emotions from live images, we can gain valuable insights into human behavior and improve our understanding of the emotional states of people in various contexts. But as this technology advances, it's crucial to carefully explore the moral and practical ramifications.

In this project we can depict the emotion of the person with the live capturing process. By this analysis method we can import the playlist that convey the same mood to the user. The dataset is retrieved from the playlist to play the random song depicting that mood of the user.

V. RESULTS AND DISCUSSIONS

As the dataset showed to the model is the primary value of the analysis, we must provide a valid group of images to the model. The dataset presented to our model helps it to increase it accuracy. At present the trained model works with an accuracy of 70% in order to detect the emotion displayed by the user. Although the calculated results of the model displayed variance in depicting the anger and sad mood, the user experience is found to be relevant to the dataset.

Human face can display 9 forms of emotions, but to understand and variate the differences between each user we need to have more live examples. Therefore, by training with more live examples and adding more datasets we can improve the accuracy of our model. Thus, this model can help to improve the user recommendation system by several music applications.

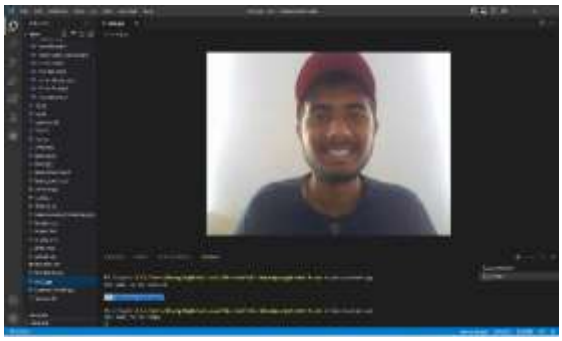

Fig. *Happy face detected*

Fig. *Happy song playing*

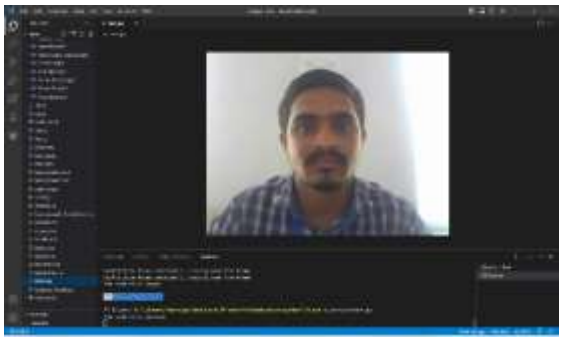

Fig. *Neutral expression detected*

Fig. *Neutral mood song played*

VI. CONCLUSIONS

This paper suggests a musical melody-generating algorithm based on facial expressions. The proposed algorithm proved successful, according to experimental results, in automating the production of musical compositions based on facial expressions, which decreased the labour and time needed to perform the activity manually. The associated costs were decreased because using a laptop or camera eliminated the need for any additional technologies, such as EEG apparatus and sensors. The algorithm's overall processing time includes both the time it takes to recognise facial emotions and the time it takes to query the meta file because face emotion recognition is not done in real time.

As a result, in terms of computational performance, the recommended method outperforms algorithms described in the present literature. The time required to recognise facial expressions is proportional to the algorithm's overall processing time. Algorithm of Haar Cascade Accurately identify the user's face. By employing a camera to capture the image, analysing the user's sentiment, and offering a personalised playlist via a more sophisticated and interactive system, it makes the end user's job easier.

VII. REFERENCES

[1] Nishanth Hegde1, Preetha S1, Sachin Bhagwat1, “Facial Expression Classifier Using Better Technique: FisherFace Algorithm”, 1Dept. of ISE, BMS College of Engineering Bangalore, India.

[2] Zhao Kun, Li Siqi, Cai Juanjuan, Wang Hui, Wang Jingling (2019). “An Emotional Symbolic Music Generation System based on LSTM Networks”, 2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference.

[3] P Kowsalya, K Sathya, T Shubhashini, S Vikram, Priyadharsni (2018). “Face Detection And Tracking Using Viola-Jones Algorithm”, Ijaerd Volume 5, Issue 04, April 2018

[4] Lopez-Rincon Omar, Starostenko Oleg, Ayala-San Gerardo Martín (2018). “Algoritmic Music Composition Based on Artificial Intelligence: A Survey”, 978-1-5386-2363- 3/18/31.00 ©2018 IEEE.

[5] Prof. Ghule Vijaykumar R., Benke Abhijeet B., Jadhav Shubham S., Joshi Swapnil (2017). “Emotion Based Music Player Using Facial Recognition”, International Journal of Innovative Research in Computer and Communication Engineering, Vol. 5, Issue 2, February 2017 Copyright to IJIRCCE DOI: 10.15680/IJIRCCE.2017.

[6] Elliot W., Eck D., Roberts A., and Abolafia D (2016). “Project Magenta: Generating longterm structure in songs and stories,”2016,https://magneta.tensorflow.org/blog/20 16/07/15/lookback-rnn-attention-rnn/

[7] Vizitiu Cristian, Nistorescu Alexandru; Marin Mihaela, Dinculescu Adrian (2015). Novel Approach to Face Expression Analysis in Determining Emotional Valence and Intensity with Benefit for Human Space Flight Studies”, Published in E-Health and Bioengineering Conference, 19-21 Nov. 2015

[8] Dureha Anukriti (2014). “An Accurate Algorithm for Generating a Music Playlist based on Facial Expressions”, International Journal of Computer Applications (0975 – 8887) Volume 100– No.9, August 2014.

[9] Ramchandra Ambika, Kumar Ravindra (2013). ”Overview Of Face Recognition System Challenges.”, International Journal of Scientific & Technology Research volume 2, issue 8, August 2013.

[10] Boulanger-Lewandowski, N., Bengio Y., Vincent P. (2012). “Modeling temporal dependencies in high-dimensional sequences: application to polyphonic music generation and transcription,” Proceedings of the 29th International Conference on Machine Learning, vol. 18, issue 13, pp. 3981-3991, 2012.

[11] Kamal Shoaib, Sayeed Dr.Farrukh, Rafeeq Mohammed (2011). “Facial Emotion Recognition for Human-Computer Interactions using hybrid feature extraction technique.”, UKSim 5th European Simposium on Computer Modelling and Simulations 2011, 196-201.

[12] Vizitiu Cristian, Nistorescu Alexandru; Marin Mihaela, Dinculescu Adrian (2015). Novel Approach to Face Expression Analysis in Determining Emotional Valence and Intensity with Benefit for Human Space Flight Studies”, Published in E-Health and Bioengineering Conference, 19-21 Nov. 2015

[13] Dureha Anukriti (2014). “An Accurate Algorithm for Generating a Music Playlist based on Facial Expressions”, International Journal of Computer Applications (0975 – 8887) Volume 100– No.9, August 2014.

[14] K. C. Tseng, B. -S. Lin, C. -M. Han and P. -S. Wang, "Emotion recognition of EEG underlying favourite music by support vector machine," 2013 1st International Conference on Orange Technologies (ICOT), Tainan, Taiwan, 2013,pp.155-158,doi:10.1109/ICOT.2013.6521181. 53

[15]Ramchandra Ambika, Kumar Ravindra (2013).” Overview Of Face Recognition System Challenges.”, International Journal of Scientific & Technology Research volume 2, issue 8, August 2013.

[16] Boulanger-Lewandowski, N., Bengio Y., Vincent P. (2012). “Modeling temporal dependencies in high-dimensional sequences: application to polyphonic music generation and transcription,” Proceedings of the 29th International Conference on Machine Learning, vol. 18, issue 13, pp. 3981-3991, 2012.

[17] Kamal Shoaib, Sayeed Dr.Farrukh, Rafeeq Mohammed (2011). “Facial Emotion Recognition for Human-Computer Interactions using hybrid feature extraction technique.”, UKSim 5th European Simposium on Computer Modelling and Simulations 2011, 196-201

[18] Amani Indunil Soysa and K. Lokuge, "ChordATune - An emotion based melody harmonizer for piano music," 2010 International Conference on Technology for Education, Mumbai, India, 2010, pp. 226-227, doi: 10.1109/T4E.2010.5550101.

[19] H. Liu, J. Hu and M. Rauterberg, "LsM: A new location and emotion aware web-based interactive music system," 2010 Digest of Technical Papers International Conference on Consumer Electronics (ICCE), Las Vegas, NV, USA, 2010, pp. 253-254, doi: 10.1109/ICCE.2010.5418750.