PAPER • **OPEN ACCESS**

Emotional Detection and Music Recommendation System based on User Facial Expression

To cite this article: S Metilda Florence and M Uma 2020 *IOP Conf. Ser.: Mater. Sci. Eng.* **912** 062007

IOP Conf. Series: Materials Science and Engineering **912** (2020) 062007 doi:10.1088/1757-899X/912/6/062007

Emotional Detection and Music Recommendation System based on User Facial Expression

S Metilda Florence[1] and M Uma[2]
[1]Department of Information Technology, SRM Institute of Science and Technology, India
[2]Department of Software Engineering, SRM Institute of Science and Technology, India

E-mail: metildam@srmist.edu.in

Abstract. It is often confusing for a person to decide which music he/she have to listen from a massive collection of existing options. There have been several suggestion frameworks available for issues like music, dining, and shopping depending upon the mood of user. The main objective of our music recommendation system is to provide suggestions to the users that fit the user's preferences. The analysis of the facial expression/user emotion may lead to understanding the current emotional or mental state of the user. Music and videos are one region where there is a significant chance to prescribe abundant choices to clients in light of their inclinations and also recorded information. It is well known that humans make use of facial expressions to express more clearly what they want to say and the context in which they meant their words. More than 60 percent of the users believe that at a certain point of time the number of songs present in their songs library is so large that they are unable to figure out the song which they have to play. By developing a recommendation system, it could assist a user to make a decision regarding which music one should listen to helping the user to reduce his/her stress levels. The user would not have to waste any time in searching or to look up for songs and the best track matching the user's mood is detected, and songs would be shown to the user according to his/her mood. The image of the user is captured with the help of a webcam. The user's picture is taken and then as per the mood/emotion of the user an appropriate song from the playlist of the user is shown matching the user's requirement.

1. Introduction

People tend to express their emotions, mainly by their facial expressions. Music has always been known to alter the mood of an individual. Capturing and recognizing the emotion being voiced by a person and displaying appropriate songs matching the one's mood and can increasingly calm the mind of a user and overall end up giving a pleasing effect. The project aims to capture the emotion expressed by a person through facial expressions. A music player is designed to capture human emotion through the web camera interface available on computing systems. The software captures the image of the user and then with the help of image segmentation and image processing techniques extracts features from the face of a target human being and tries to detect the emotion that the person is trying to express. The project aims to lighten the mood of the user, by playing songs that match the requirements of the user by capturing the image of the user. Since ancient times the best form of expression analysis known to humankind is facial expression recognition. The best possible way in which people tend to analyze or conclude the emotion or the feeling or the thoughts that another

person is trying to express is by facial expression. In some cases, mood alteration may also help in overcoming situations like depression and sadness. With the aid of expression analysis, many health risks can be avoided, and also there can be steps taken that help brings the mood of a user to a better stage.

2. Literature Survey

Renuka R Londhe et al. [1] proposed a paper which focused on the study of changes in the curvatures of the face and the intensities of the corresponding pixels. The author used Artificial Neural Networks(ANN), which was used to classify the emotions. The author also proposed various approaches for a playlist. Zheng et al. [2]proposed two significant categories for facial feature extraction, which included Appearance-based feature extraction and geometric based feature extraction, which included extraction of some essential points of the face such as mouth, eyes, and eyebrows.Nikhil et al. [3].determines the mindset of the user by using facial expression Humans often express their feeling by their expressions, hand gestures, and by raising the voice of tone but mostly humans express their feelings by their face. Emotion-based music player reduces the time complexity of the user. Generally, people have a large number of songs on their playlist. Playing songs randomly does not satisfy the mood of the user. This system helps user to play songs automatically according to their mood. The image of the user is captured by the web camera, and the images are saved. The images are first converted from RGB to binary format. This process of representing the data is called a feature-point detection method. This process can also be done by using Haar Cascade technology provided by Open CV. The music player is developed by using a java program. It manages the database and plays the song according to the mood of the user.Z. Zeng et al.[5] researched various advances in human affect recognition. He focused on various approaches that can handle audio and/or visual recordings of displays of affective states. The paper provides a detailed review of audio/visual computing methods. The effect is described as a prototype of emotion categories which include happiness, sadness, fear, anger, disgust, and surprise. This paper focused on discussing the challenges in computing methods for the development of automatic, spontaneous affect recognizer, which helped in emotion detection. It also identified some problems that have been missed or avoided in uni-modal posed emotion recognition. Parul Tambe et al. [7] proposed an idea which automated the interactions between the users and music player, which learned all the preferences, emotions and activities of a user and gave song selection as a result. The various facial expressions of users were recorded by the device to determine the emotion of the user to predict the genre of the music. Jayshree Jha et al. [11].proposed an emotion-based music player using image processing This showed how various algorithms and techniques that were suggested by different authors in their research could be used for connecting the music player along with human emotions. It has thus helped in reducing the efforts of user in creating and managing playlist and providing an excellent experience to the music listeners by bringing them the most suitable song according to the user's his/her current expression. Anukritine et al. [18] came up with an algorithm that gives a list of songs from the user's playlist in accordance with the user's emotion. The algorithm which was designed was focused on having less computational time and also thus reduces the cost included in using various hardware. The main idea was to segregate the emotions into five categories i.e., Joy, sad, anger, surprise and fear also provided a highly accurate audio information retrieval approach that extracted relevant information from an audio signal in less time. Aditya et al. [19] developed an android application which acts as a customized music player for a user using image processing to analyze and present songs to user according to the user's mood. The application was developed using Eclipse and also OpenCV to implement facial recognition algorithms. This paper also showed comparison between various algorithms used in facial detection. The images of the user were captured using the front camera of the mobile device. It aimed to provide satisfaction to music lovers by extracting their emotions. A. Habibzad et al. [20] proposed a new algorithm is proposed to recognize the facial emotion, which included three stages: pre-processing, feature extraction, and classification. The first part describes various stages in image processing include preprocessing, filtering used to extract various facial features. The second part optimized the eye and

lip ellipse characteristics, and in the third part, the eye and lip optimal parameters were used to classify the emotions. The obtained results showed that the speed of facial recognition was far better than other usual approaches. Prof. Nutan Deshmukh et al. [21] focused on creating a system that fetches the emotion of the user using a camera and then automates the result using the emotion detection algorithm. This algorithm captures the mood of the user after every decided interval of time as the mood of the user may not be the same after some time; it may or may not change. The proposed algorithm on an average calculated estimation takes around 0.95-1.05 sec to generate an emotion-based music system, which was better than previous existing algorithms and reduces the cost of designing. Chang Liuet al[24] described a system that makes use of Brain-Computer Interfaces, also called as BCI. BCI makes use of devices to send signals to the processing systems. EEG hardware is used in to monitor the person's cognitive state of mind. The drawback of the scheme is that they require the input from the user's brain continuously to perform the classification. An algorithm based on MID is used to continuously monitor and process the signals received from the brain of the user and use these signals to actively monitor and generate emotions that the user is currently experiencing. Swati Vaid etal[23] reviewed EEG - Electroencephalography (EEG) is a form of medical science that records the electrical activity from the neurons of brain cells. The electrical activity of the neurons from within the cells of the brain is registered. Based on the recorded activity of the neurons an approximation is made, and the emotion of the person is estimated from that analysis. This method mentioned above, although serves the purpose of getting the activity of brain cells but fails to serve the purpose of portability and economics.

3. Data collection and Data source

A survey was collected from users based on 3 parameters which are, 1. What type of songs would they want to listen to when they are happy? 2. What type of songs would they want to listen to when they are sad? 3. What type of songs would they want to listen to when they are angry?

CK+ Dataset - The dataset that is used to train the classifier is Cohn Kanade Extended dataset. The dataset consists of 593 Facial Action Coded Sequences from 123 subjects. The labeling done is to tell us about the expression that is being expressed by the subject. There are a series of images that start from the neutral expression of the target and end with the extreme emotion expressed by the subject. In our analysis and training network, the first image and the last image are used. The other images are not used as such to train the network.

HELEN Dataset - The dataset contains around 200 images are used for the training of the classifier. Along with the images, 164 landmarks positions are provided for each and every image present in the dataset in the form of a .txt file. These coordinates are extracted from the text files and used in the system to generate the .xml file. This XML file is further used to train the classifier. This trained classifier is used to predict the position of the landmarks in the other set of unknown images.

4. Proposed Architecture

The proposed system can detect the facial expressions of the user and based on his/her facial expressions extract the facial landmarks, which would then be classified to get a particular emotion of the user. Once the emotion has been classified the songs matching the user's emotions would be shown to the user.

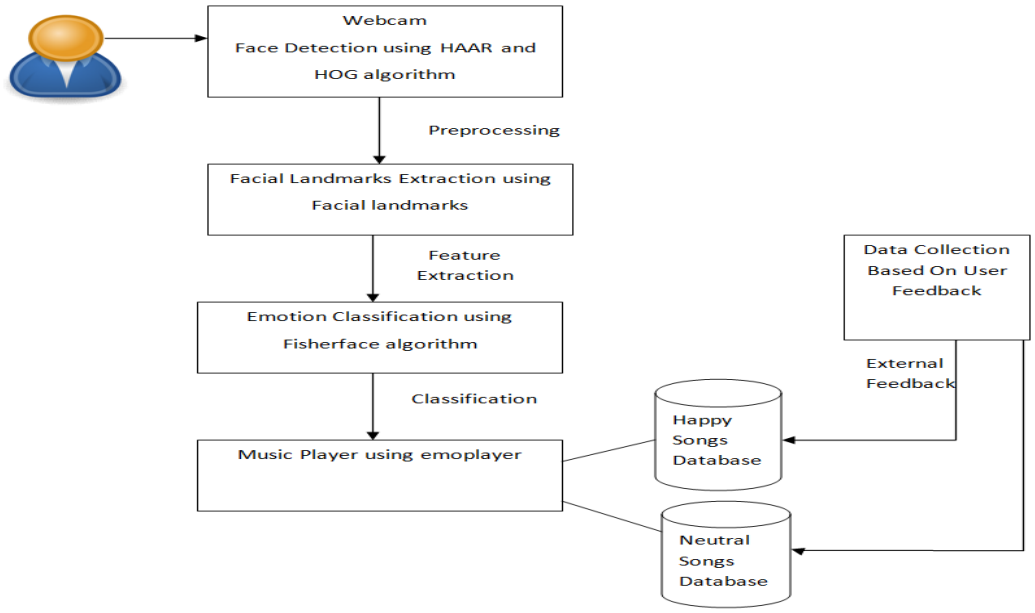

Figure 1. Architecture of Music Recommendation System Based On User Facial Expression.

4.1. Methodologies

- **Emotion Extraction Module** -The image of the user is captured with the help of a camera/webcam. Once the picture captured, the frame of the captured image from webcam feed is converted to a grayscale image to improve the performance of the classifier, which is used to identify the face present in the picture. Once the conversion is complete, the image is sent to the classifier algorithm which, with the help of feature extraction techniques can extract the face from the frame of the web camera feed. From the extracted face, individual features are obtained and are sent to the trained network to detect the emotion expressed by the user. These images will be used to train the classifier so that when a completely new and unknown set of images is presented to the classifier, it is able to extract the position of facial landmarks from those images based on the knowledge that it had already acquired from the training set and return the coordinates of the new facial landmarks that it detected. The network is trained with the help of CK extensive data set. This is used to identify the emotion being voiced by the user
- **Audio Extraction Module** - After the emotion of the user is extracted the music/audio based on the emotion voiced by the user is displayed to the user, a list of songs based on the emotion is displayed, and the user can listen to any song he/she would like to. Based on the regularity that the user would listen to the songs are displayed in that order. This module is developed using web technologies like PHP, MySQL, HTML, CSS, JAVASCRIPT.
- **Emotion - Audio Integration Module** - The emotions which are extracted for the songs are stored, and the songs based on the emotion are displayed on the web page built using PHP and MySQL. For example, if the emotion or the facial feature is categorized under happy, then songs from the happy database are displayed to the user.

5. Experiment Results & Analysis

5.1. Results

This study proposes a music recommendation system which extracts the image of the user, which is captured with the help of a camera attached to the computing platform. Once the picture has been

captured, the captured frame of the image from webcam feed is then being converted to a grayscale image to improve the performance of the classifier that is used to identify the face present in the picture. Once the conversion is complete, the image is sent to the classifier algorithm which, with the help of feature extraction techniques is able to extract the face from the frame of the web camera feed. Once the face is extracted individual features from the face is extracted and is sent to the trained network to detect the emotion expressed by the user. A classifier that is used to detect or obtain the facial landmarks from the face of the user is trained on HELEN dataset. HELEN dataset contains more than 2000 images. These images will be used to train the classifier so that when a completely new and unknown set of images is presented to the classifier, it is able to extract the position of facial landmarks from those images based on the knowledge that it had already acquired from the training set and return the coordinates of the new facial landmarks that it detected. The network is trained with the help of CK extensive data set. This is used to identify the emotion being voiced by the user. Once this has been detected, an appropriate song is selected by the music player that would best match the mood of the user. The overall idea behind making the system is to enhance the experience of the user and ultimately relieve some stress or lighten the mood of the user. The user does not have to waste any time in searching or to look up for songs and the best track matching the user's mood is detected and played automatically by the music player. The image of the user is captured with the help of a webcam. The user's picture is taken and then as per the mood/emotion of the user an appropriate song from the playlist of the user is played matching the user's requirement.

The system has successfully been able to capture the emotion of a user. It has been tested in a real-time environment for this predicate. It has to be, however, tested in different lighting conditions to determine the robustness of the developed system. The system has also been able to grab the new images of the user and appropriately update its classifier and training dataset. The system was designed using the facial landmarks scheme and is tested under various scenarios for the result that would be obtained. It is seen that the classifier has an accuracy of more than 80 percent for most of the test cases, which is pretty good accuracy in terms of emotion classification. It can also be seen that the classifier can accurately predict the expression of the user in a real-time scenario when tested live for a user.

5.2. Results Analysis

5.2.1. Experiment Results- Instructions Explained to the User. In this scenario the users were given instructions as to what is to be done to perform the prediction of the emotion expressed which provided the following results. Sometimes in cases where the inner emotion is sad and facial expression is happy it resulted in a fail case. The values are given in Table 1 and the result is shown in Figure 2.

Table 1. Instructions Explained to the User.

User	Emotion	Facial Expression	Accuracy
1	Happy	Happy	100
2	Sad	Happy	0
3	Happy	Happy	100
4	Sad	Sad	100

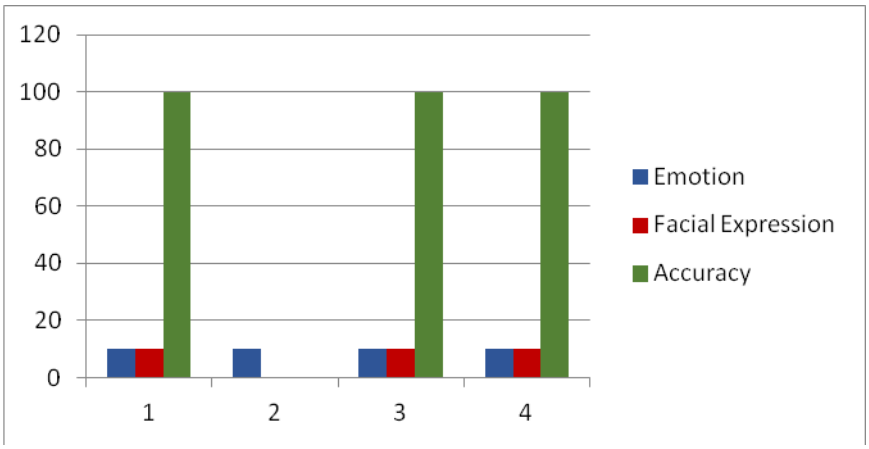

Figure 2. Experiment Results- Instructions Explained to the User.

5.2.2. Experiment Results - Instructions not given to the User. In this scenario the users were not given any instructions as to what is to be do and thus the inner emotions or the emotions recognized failed, there were also cases where in the emotion matched with the facial expressions of the user. The values are given in Table 2 and the result is shown in Figure 3.

Table 2. Instructions not explained to the User.

User	Emotion	Facial Expression	Accuracy
1	Happy	Sad	0
2	Sad	Happy	0
3	Sad	Sad	100
4	Happy	Sad	0

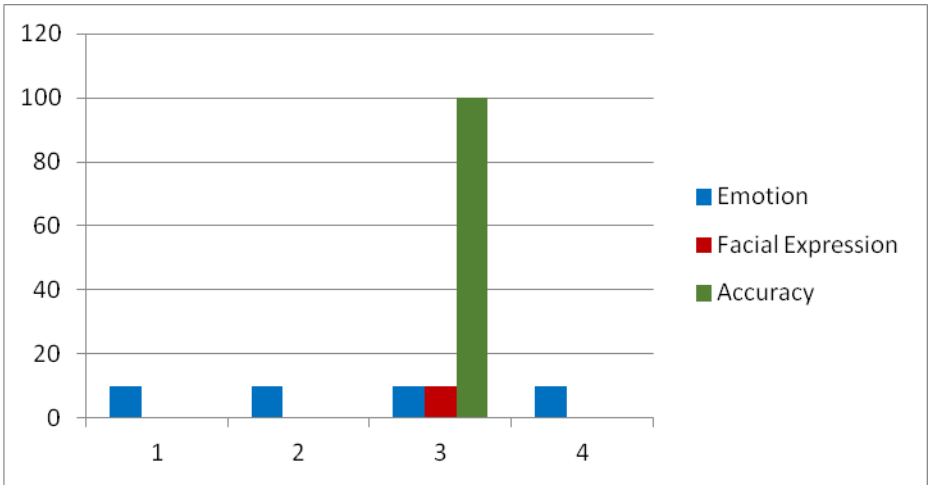

Figure 3. Experiment Results- Instructions not explained to the User.

6. Conclusion & Future Work

Emotion recognition using facial expressions is one of the important topics of research and has gathered much attention in the past. It can be seen that the problem of emotion recognition with the help of image processing algorithms has been increasing day by day. Researchers are continuously working on ways to resolve this by the use of different kinds of features and image processing methods. The applications of image processing algorithms in the field of both medical science and

human science are of vast importance. There are continuously new ways and methods being developed that make use of image processing algorithms to extract the emotion of the user and make use of the extracted emotion to treat the user. Emotion recognition has gained a lot of importance in all aspects of life and if a robust algorithm implemented which can accurately classify the emotions of the person, then a great deal of advancement in the industry can be achieved with the help of this. The system has successfully been able to capture the emotion of a user. It has been tested in a real-time environment for this predicate. However, it has to be tested in different lighting conditions to determine the robustness of the developed system. The system has also been able to grab the new images of the user and appropriately update its classifier and training dataset. The system was designed using the facial landmarks scheme and was tested under various scenarios for the result that would be obtained. It is seen that the classifier has an accuracy of more than 80 percent for most of the test cases, which is pretty good accuracy in terms of emotion classification. It can also be seen that the classifier can accurately predict the expression of the user in a real-time scenario when tested live for a user.

Limitations

- The system still is not able to record all the emotions correctly due to the less availability of the images in the image dataset being used.
- The image that is fed into the classifier should be taken in a well-lit atmosphere for the classifier to give accurate results.
- The quality of the image should be at least higher than 320p for the classifier to predict the emotion of the user accurately.

Future Work

- Reduce the time required to train the classifier
- Use of EEG signals to make the software even more optimized and to detect the exact mood /emotion of the user.

7. References

[1] Londhe RR and Pawar DV 2012 Analysis of facial expression and recognition based on statistical approach *International Journal of Soft Computing and Engineering* **2**

[2] Kabani H, Khan S, Khan O and Tadvi S 2015 Emotion based music player *International Journal of Engineering Research and General Science* **3** 750-6

[3] Gupte A, Naganarayanan A and Krishnan M Emotion Based Music Player-XBeats *International Journal of Advanced Engineering Research and Science* **3** 236854

[4] Hadid A, Pietikäinen M and Li SZ 2007 Learning personal specific facial dynamics for face recognition from videos *International Workshop on Analysis and Modeling of Faces and Gestures* pp1-15 Springer Berlin Heidelberg

[5] Zeng Z, Pantic M, Roisman GI and Huang TS 2008 A survey of affect recognition methods Audio, visual, and spontaneous expressions *IEEE transactions on pattern analysis and machine intelligence* **31** 39-58

[6] Patel AR, Vollal A, Kadam PB, Yadav S and Samant RM 2016 MoodyPlayer a mood based music player *Int. J. Comput. Appl*. **141** 0975-8887

[7] ParulTambe, YashBagadia, Taher Khalil and Noor UlAin Shaikh 2015 Advanced Music Player with Integrated Face Recognition Mechanism *International Journal of Advanced Research in Computer Science and Software Engineering* **5**

[8] Lucey P, Cohn JF, Kanade T, Saragih J, Ambadar Z and Matthews I 2010 The extended cohn-kanade dataset (ck+) A complete dataset for action unit and emotion-specified expression In 2010 *ieee computer society conference on computer vision and pattern recognition-workshops* 94-101 IEEE

[9] Kanade T, Cohn JF and Tian Y 2000 Comprehensive database for facial expression analysis *InProceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition* 46-53 IEEE
[10] Luoh L, Huang CC and Liu HY 2010 Image processing based emotion recognition In2010 *International Conference on System Science and Engineering* 491-494 IEEE
[11] Vivek JD, Gokilavani A, Kavitha S, Lakshmanan S and Karthik S 2017 A novel emotion recognition based mind and soul-relaxing system *In2017 International Conference on Innovations in Information, Embedded and Communication Systems* 1-5 IEEE
[12] Jyoti Rani and Kanwal Garg 2014 Emotion Detection Using Facial Expressions A Review *International Journal of Advanced Research in Computer Science and Software Engineering* **4**
[13] Joshi A and Kaur R 2013 A Study of speech emotion recognition methods *Int. J. Comput. Sci. Mob. Comput*. **2** 28-31
[14] Shoaib M, Hussain I, Mirza HT and Tayyab M 2017 The role of information and innovative technology for rehabilitation of children with autism a systematic literature review In2017 *17th International Conference on Computational Science and Its Applications* 1-10 IEEE
[15] Dubey M and Singh L 2016Automatic emotion recognition using facial expression a review *International Research Journal of Engineering and Technology (IRJET)* **3** 488-92
[16] Anwar S, Milanova M, Bigazzi A, Bocchi L and Guazzini A 2016 Real time intention recognition *In IECON 2016-42nd Annual Conference of the IEEE Industrial Electronics Society* 1021-1024 IEEE
[17] Rázuri JG, Sundgren D, Rahmani R, Moran A, Bonet I and Larsson A 2015 Speech emotion recognition in emotional feedbackfor human-robot interaction *International Journal of Advanced Research in Artificial Intelligence* **4** 20-7
[18] Dureha A 2014 An accurate algorithm for generating a music playlist based on facial expressions *International Journal of Computer Applications* **100** 33-9
[19] Gupte A, Naganarayanan A and Krishnan M Emotion Based Music Player-XBeats *International Journal of Advanced Engineering Research and Science* **3** 236854
[20] A. habibzad, ninavin and Mir kamalMirnia A new algorithm to classify face emotions through eye and lip feature by using particle swarm optimization 2012 *4th International Conference on Computer Modeling and Simulation*
[21] Rabashette MD, Tale MR, Hinge MA, Padale MK, Chavan MR and Deshmukh N Emotion Based Music System
[22] Lucey P, Cohn JF, Kanade T, Saragih J, Ambadar Z and Matthews I 2010 The extended cohn-kanade dataset (ck+) A complete dataset for action unit and emotion-specified expression In 2010 *ieee computer society conference on computer vision and pattern recognition-workshops* 94-101 IEEE
[23] Vaid S, Singh P and Kaur C 2015 EEG signal analysis for BCI interface: A review In2015 *fifth international conference on advanced computing & communication technologies* 143-147 IEEE
[24] Liu C, Xie S, Xie X, Duan X, Wang W and Obermayer K 2018 Design of a video feedback SSVEP-BCI system for car control based on improved MUSIC method In2018 *6th International Conference on Brain-Computer Interface (BCI)* 1-4 IEEE
[25] Uma M and Sheela T 2018 *IOP Conf. Ser.: Mater. Sci. Eng.* **402** 012182
[26] Uma M and Sheela T 2017 Analysis of Collaborative Brain Computer Interface (BCI) based Personalized GUI for Differently Abled *Intelligent Automation & Soft Computing* **29**