International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 10 Issue: 04 | Apr 2023 www.irjet.net p-ISSN: 2395-0072

Emotion Based Music Player System

Salvina Desai[1], Vrushali Paul[2], Sayali Jadhav[3], Pranita Gode[4], Roshani Bhaskarwar[5]

Department of Information Technology, Datta Meghe College Of Engineering, Airoli, Navi Mumbai, Maharashtra, India

---

Abstract - *This paper proposes the implementation of an intelligent agent that segregates songs and plays them according to the user's current mood. The music that best matches the emotion is recommended to the user as a playlist. Face emotion recognition is a form of image processing. Facial emotion recognition is the process of converting the movements of a person's face into a digital database using various image processing techniques. Facial Emotion Recognition recognises the face's emotions. The collection of songs is based on the emotions conveyed by each song and then suggests an appropriate playlist. The user's music collection is initially clustered based on the emotion the song conveys. This is calculated taking into consideration the lyrics of the song as well as the melody. Every time the user wishes to generate a mood-based playlist, the user takes a picture of themselves at that instant. This image is subjected to facial detection and emotion recognition techniques, recognising the emotion of the user.*

***Key Words*: music classification, music recommendation, emotion recognition, intelligent music player, face recognition, feature extraction**

1. INTRODUCTION

Emotion recognition is a feature of artificial intelligence that is becoming more applicable for robotically performing various processes that are relatively more exhausting to perform manually. The human face is an essential part of the human body, mostly when it comes to extracting a person's emotional state and behaviour according to the situation [1]. Recognizing a person's mood or state of mind based on the emotions they show is an important part of making systematic decisions that are best suited to the person in question for a diversity of applications.

In day to day life, each person faces a lot of problems, and the best helper for all the stress, anxiety, tension, and worry that is encountered is music [2]. Music plays a very vital role in building up and enhancing the life of every individual because it is an important medium of entertainment for music lovers and listeners. In today's world, with ever-increasing advancement in the field of technology and multimedia, several music players were developed with functions like fast reverse, forward, variable playback speed (speeding up or slowing down the original speed of audio), streaming playback, volume modulation, genre classification, etc. Although these functions satisfy the basic requirements of the user, the user still has to manually scroll through the playlist and choose songs based on his current mood and behavior.

Imagine yourself in a world where humans interact with computers. You are sitting in front of your personal computer, which can listen, talk. It has the ability to gather information about you and interact with you through special techniques like facial recognition, speech recognition, etc. It can even understand your emotions at the touch of a mouse. It verifies your identity, feels your presence, and starts interacting with you. The vital part of hearing the song has to be done in a facilitated way; that is, the player has to be able to play the song in accordance with the person's mood. People tend to reveal their emotions, mainly through facial expressions. Capturing emotions, recognising the emotions of a person, and displaying suitable songs corresponding to his mood can help to calm the user's mind, which has a satisfying effect on the user.

This project aims to capture the emotions expressed by a user's facial expressions, and the music player is structured to capture a person's emotions through a webcam interface available on a computer. The software captures the user's image, and subsequently, using image segmentation and image processing techniques, it extracts features from the face of a person and tries to reveal the emotions that person is trying to express. The aim of the project is to lighten the user's mood by playing songs that match user requests by capturing the user's emotion through the image. Since ancient times, the best form of expression analysis known to mankind has been facial expression recognition.

It will also help in the entertainment field, for the purpose of providing recommendations to an individual person based on their current mood. We study this from the perspective of providing a person with customized music recommendations based on their state of mind as detected from their facial expressions [3]. Most music connoisseurs have extensive music collections that are often sorted only based on parameters such as artist, album, genre, and number of times played. However, this often leaves the users with the arduous task of making mood-based playlists, finding the music based on the emotion conveyed by the songs—something that is much more essential to the listening experience than it often appears to be. This task only increases in complexity with larger music collections and consumes a lot of time; automating the process would save many users time and the effort spent in doing the same manually selection of songs, while improving their overall

experience and allowing for a better enjoyment of the music. It recognises the emotion on the user's face and plays songs according to that emotion.

2. RELATED WORK

R. Ramanathan et.al have proposed the first part of the emotion-based music player system is emotion recognition. The software captures the emotion of a person in the image captured by the webcam using various image processing and segmentation techniques. bit extracts features from the face of a person. The available dataset is used after training for classification, clustering, and emotion recognition for faces. The system's next step entails categorising music and assigning labels to each song in accordance with the emotions it conveys. Music and emotion have been the subject of research. Feature extraction from each song based on the best features to extract is the first step in recognition, starting with the generation of relevant datasets to get the most accurate results [1].

Charles Darwin was the first scientist to recognise that facial expression is one of the most powerful and immediate means for human beings to communicate their emotions, intentions, and opinions to each other [4]. Rosalind Picard (1997) describes why emotions are important to the computing community. There are two aspects to effective computing: giving the computer the ability to detect emotions and giving it the ability to express emotions. Not only are emotions crucial for rational decision-making, as Picard describes, but emotion detection is an important step in an adaptive computer system. An adaptive, smart computer system has been driving our efforts to detect a person's emotional state. An important element of incorporating emotion into computing is improving the productivity of the computer user.

In 2011 [5], Ligang Zhang and Dian T developed a facial emotion recognition system (FER). They used a dynamic 3D Gabor feature approach and obtained the highest correct recognition rate (CRR) on the JAFFE database, and FER is among the top performers on the Cohn-Kanade (CK) database using the same approach. They attested to the effectiveness of the proposed approach through recognition performance, computational time, and comparison with the state-of-the-art performance. And concluded that patch-based Gabor features show a better performance over point-based Gabor features in terms of extracting regional features, keeping the position information, achieving a better recognition performance, and requiring a smaller number.

As per the survey done by R. A. Patil, Vineet Sahula, and A. S. Mandal for CEERI Pilani on expression recognition, the problem is divided into three subproblems: face detection, feature extraction, and facial expression classification. Most of the existing systems assume that the presence of the face in a scene is ensured. Most of the systems deal with only feature extraction and classification, assuming that the face is already detected [6].

The algorithms that we have selected for detecting emotions are from the papers Image Edge Detection Algorithm Based on an Improved Canny Operator of 2012 and Rapid Object Detection Using a Boosted Cascade of Simple Features by Viola and Jones. In "Image Edge Detection Algorithm Based on an Improved Canny Operator," an improved canny edge detection algorithm is proposed [7]. Because the traditional canny algorithm has difficulty treating images that contain salt and pepper noise and because it does not have the adaptive ability to adjust for the variance of the Gaussian filtering, a new canny algorithm is presented in this paper, in which open-close filtering is used instead of Gaussian filtering. In this paper, the traditional canny operator is improved by using morphology filtering to preprocess the noise image. The final edge image can effectively reduce the influence of noise, keep the edge strength and more complete details, and get a more satisfactory subjective result. And by using objective evaluation standards, compared with the traditional Canny operator, information entropy, average gradient, peak signal-to-noise ratio, correlation coefficient, and distortion degree have also increased significantly. So, the new algorithm is an effective and practical method of edge detection. Automatic face recognition is all about extracting those meaningful features from an image, putting them into a useful representation, and performing some kind of classification on them. Face recognition based on the geometric features of a face is probably the most intuitive approach to face recognition [7].

3. DESIGN & METHODOLOGY

The user's image is captured using a camera or webcam. Once the image is captured, the frame of the captured image from the webcam source is converted to a gray-scale image to improve the performance of the classifier, which is used to identify the face in the image. Once the conversion is complete, the image is sent to a classification algorithm that can be extracted using feature extraction techniques from the webcam feed frame. From the extracted face, individual features are obtained and sent to a trained network to detect the emotions 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 it, it can extract the location of facial landmarks from these images using the knowledge it gained from the training set and return the coordinates of the new facial landmarks it detected. The network is trained with the help of an extensive data set. This is used to identify the emotion it expresses to the user.

Now that detected face can use image for face recognition. However, if it simply performs face recognition directly on a normal photo image, it would probably get less than 10% accuracy! The system was designed to be a better, more efficient, and less space-consuming product that users could effectively apply and use, that could be tested, and that was simple to configure. The figure 1 shows the components that the project uses to perform the required work and tasks and structure in the project model.

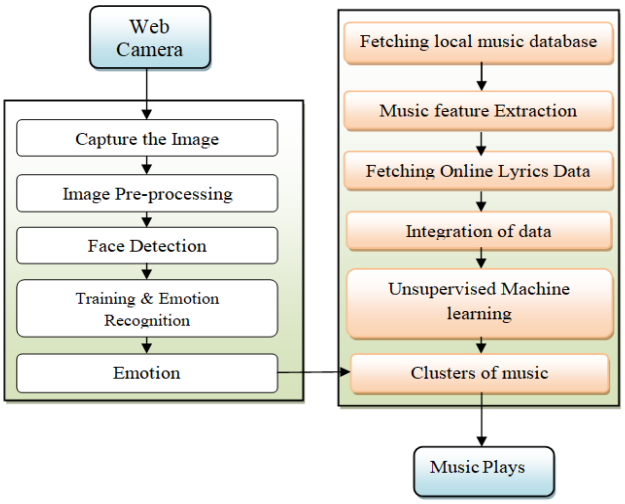

Fig -1: General block diagram of the System design

As shown in the above figure, It is extremely important to apply various image pre-processing techniques to standardise the images that supply a face recognition system. Most face recognition algorithms are extremely sensitive to lighting conditions, so if they were trained to recognise a person when they are in a bright room, they probably won't recognise them in a dark room, etc. This problem is referred to as "lighting dependent," and there are also many other issues, such as the fact that the face should also be in a very consistent position within the images (such as the eyes being in the same pixel coordinates), consistent size, rotation angle, hair and makeup, emotion (smiling, angry, etc.), and position of lights (to the left or above, etc.). This is why it is so important to use good image pre-processing filters before applying face recognition it is also do things like removing the pixels around the face that aren't used, such as with an elliptical mask to only show the inner face region and not the hair and image background, since they change more than the face does. For simplicity, the face recognition system will show Eigen faces and grayscale images [8]. So it will show how to easily convert colour images into grayscale and then easily apply histogram equalisation as a very simple method of automatically standardising the brightness and contrast of your facial images. For better results, it can use colour face recognition (ideally with a colour histogram fitting in HSV or another colour space instead of RGB) or apply more processing stages such as edge enhancement, contour detection, motion detection, etc. Also, this code is resizing images to a standard size, but this might change the aspect ratio of the face.

Data Flow Diagram

A data flow diagram (DFD) is a graphical representation of the flow of data through an information system. A data flow diagram can also be used for the visualization of structured data processing design. It is common practice for a designer to first draw a context-level DFD, which shows the interaction between the system and outside entities. This context-level DFD is then used to show more detail about the system being model.

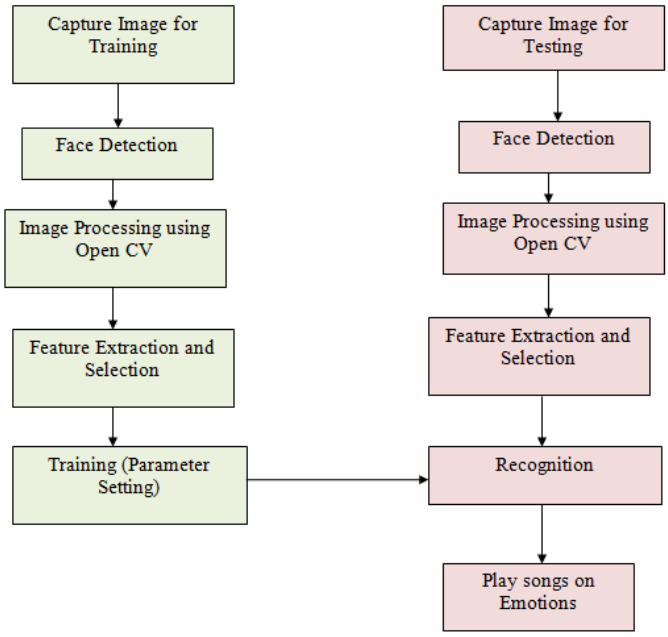

Fig -2: DFD Diagram of the system

A. Face Capturing

We used open-source computer vision (OpenCV), a library of programming functions aimed mainly at real-time computer vision. It's usually used for real-time computer vision, so it's easier to combine with other libraries that can also use NumPy [9]. When the first process starts, the stream from the camera is accessed, and about 10 photos are taken for further processing and emotion recognition. We use an algorithm to categorise photos, and for that we need a lot of positive images that only contain images of people's faces and also negative images that only contain images of people without faces. Instruct the classifier. The model is built using classified photos.

B. Face Detection

The principal component analysis (PCA) method is used to reduce the face space dimensions. Following that, linear discriminant analysis (LDA) method is used to obtain image feature characteristics. We use this method specifically because it maximises the training process's classification between classes. While using the minimal Euclidean approach for matching faces, this algorithm aids in picture recognition processing and helps us categorise user expressions that suggest emotions.

B.1 Dataset

There will be one dataset for images, which we have used from the Kaggle collection for our project. The input images to the system will be real-time images. The dataset used here for images contains images of different facial expressions. Every emotion captured in an image is represented by at least one image in the data file. Then the data set will be divided into the moods specified on the label of the data set. It will classify moods accordingly by recognizing a person's facial expressions, like happy, sad, surprised, afraid, angry, sad, and neutral.

C. Facial Emotion Recognition

A Python script is used to fetch images containing faces along with their emotional descriptor values. The images are contrast enhanced by contrast-limited adaptive histogram equalisation and converted to grayscale in order to maintain uniformity and increase the effectiveness of the classifiers. A cascade classifier, trained with face images, is used for face detection, where the image is split into fixed-size windows. Each window is passed to the cascade classifiers and is accepted if it passes through all the classifiers; otherwise, it is rejected. The detected faces are then used to train the face, which works on reducing variance between classes. Fisher's face recognition method proves to be efficient as it works better with additional features such as spectacles and facial hair. It is also relatively invariant to lighting. A picture is taken during the runtime of the application, which, after pre-processing, is predicted to belong to one of the emotion classes by the fisher face classifier. The model also permits the user to customize the model in order to reduce the variance within the classes further, initially or periodically, such that the only variance would be that of emotion changes [10].

D. Feature Extraction

The features considered while detecting emotion can be static, dynamic, point based geometric, or region based appearance. To obtain real-time performance and to reduce time complexity, for the intent of expression recognition, only eyes and mouth are considered. The combination of two features is adequate to convey emotions accurately. Finally, In order to identify and segregate feature points on the face, a point detection algorithm is used [2].

- Eye Extraction: The eyes display strong vertical edges (horizontal transitions) due to its iris and eye white. In order to find the Y coordinate of the eyes, vertical edges from the horizontal projection of the image is obtained through the use of Sobel mask [14].
- Eyebrow Extraction: Two rectangular regions in the edge image which lie directly above each of the eye regions are selected as the eyebrow regions. The edge images of these two areas are obtained for further refinement. Now Sobel method was used in obtaining the edge image as more images can be detected when compared to Robert's method. These obtained edge images are then dilated and the holes are filled. The result edge images are used in refining the eyebrow regions.
- Mouth Extraction: The points in the top region, bottom region, right corner points and left corner points of the mouth are all extracted and the centroid of the mouth is calculated.

E. Music Recommendation

The emotion detected from the image processing is given as input to the clusters of music to select a specific cluster. In order to avoid interfacing with a music app or music module, which would involve extra installation, the support from the operating system is used instead to play the music file. The playlist selected by the clusters is played by creating a forked sub process that returns control back to the python script on completion of its execution so that other songs can be played. This makes the programme play music on any system, regardless of its music player.

F. Haar cascade classifiers

Multiple images are captured from a web camera. To predict the emotion accurately, we might want to have more than one facial image. Blurred images can be an error source (especially in low light conditions) and hence, the multiple images are averaged to get an image devoid of any blur. Histogram equalization is an image processing technique used to enhance the contrast of the image by normalizing the image throughout its range. This image is then cropped and converted to greyscale so that only the foreground of the image remains, thereby reducing any ambiguity. A Haar classifier is used for face detection where the classifier is trained with pre-defined varying face data which enables it to detect different faces accurately [13]. The core basis for Haar classifier object detection is Haar-like features. These features, rather than using the intensity values of a pixel, use the change in contrast values between adjacent rectangular groups of pixels [11]. The contrast variances between the pixel groups are used to determine relative light and dark areas. Two or three adjacent groups A Haar-like feature is formed by a relative contrast variance. Haar-like features are used to detect an image. Haar features can easily be scaled by increasing or decreasing the size of the pixel group being examined. This allows features to be used to detect objects of various sizes [8]. Each Haar-like feature consists of two or three jointed "black" and "white" rectangles: A collection of basic Haar-like characteristics.

The value of a Haar-like feature is the difference between the sums of the pixel grey level values within the black and white rectangular regions:

$F(x)=Sum_{black\ rectangle}$ *(pixel gray level)* $- Sum_{white\ rectangle}$ *(pixel gray level)*

Compared with raw pixel values, Haar-like features can reduce or increase the in-class or out-of-class variability, and this makes classification easier [9].

4. IMPLEMENTATION & RESULTS

Following are the implementation and result screenshots of our project.

Home Page:

The homepage is the index page of the music recommendation system. Registration and login options are available on the homepage. By clicking on this option, users can register and login to their accounts, as shown in Figure 3.

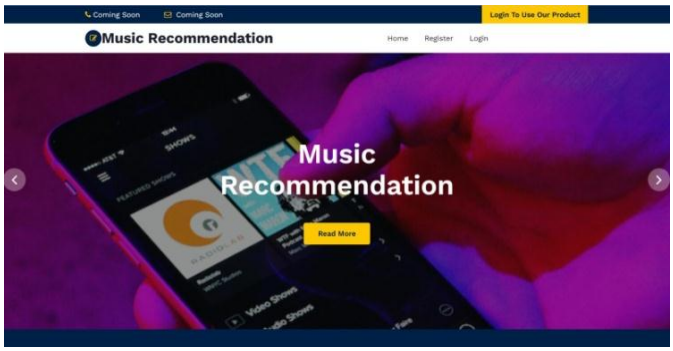

Fig -3: Home page

Registration Page:

This registration page is for users who wish to register themselves as users. They can successfully register after filling out all of the required information, as shown in Figure 4.

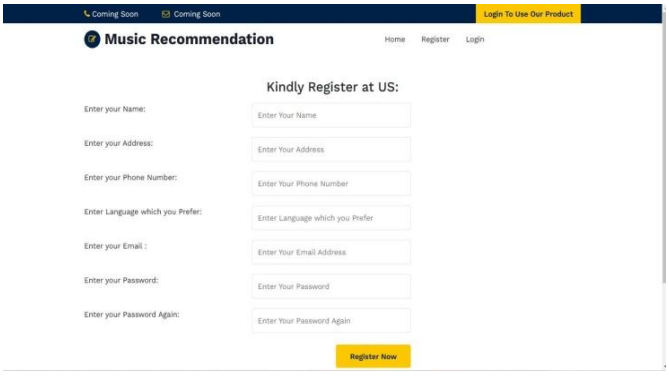

Fig -4: Registration page

Login Page:

As shown in Figure 5, a user can log in to their profile account after filling in the right information, such as their email and password. The information is sent to the database to check for a match. If no match is found, the customer remains on the same page; otherwise, he is directed to their profile page.

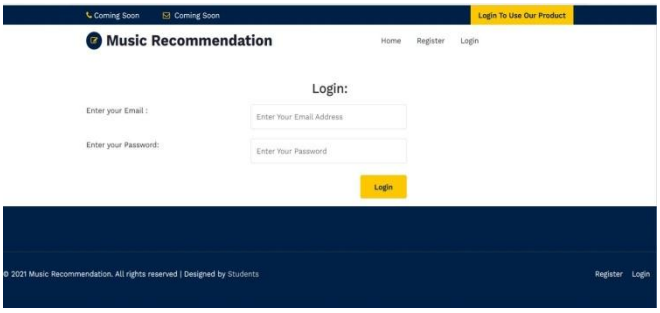

Fig -5: Login page

Upload/Capture Image:

When a user logs in, they are directed to a page where they can capture images and upload photos, as shown in Figure 6.

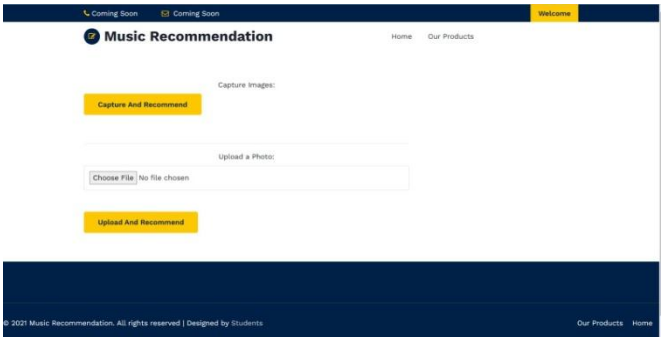

Fig -6: Image capturing page

Emotion Recognition:

As shown in Figure 7, once the camera is open, it will capture the image and identify the emotions.

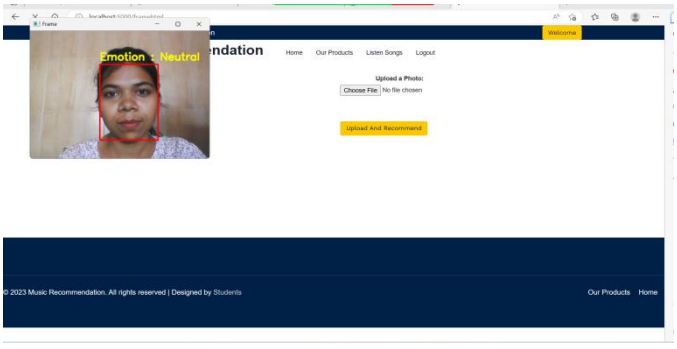

Fig -7: Image capturing page

Search song:

The user can also search for the artist, song, and album of their choice, as shown in Figure 8.

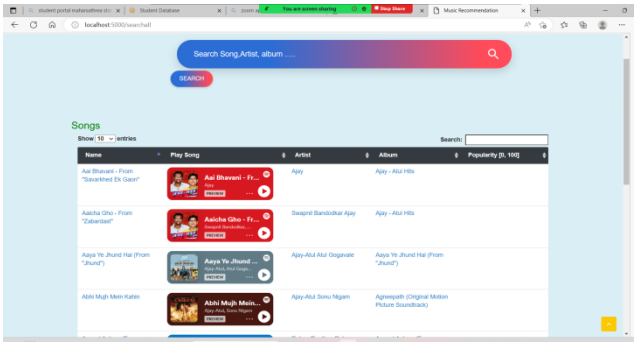

Fig -8: Song searching page

Music Recommendation Page:

Once an image is captured or a photo is uploaded, emotion is detected in it, and a particular playlist of songs is recommended to the user, as shown in Figure 9.

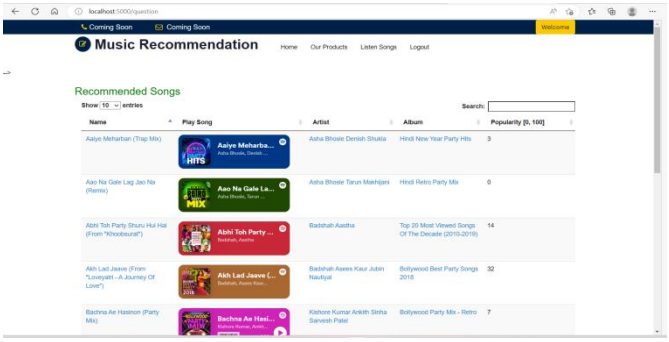

Fig -9: Track recommendation page

5. CONCLUSION

The aim of this paper is to explore the area of automatic facial expression recognition for the implementation of an emotion-based music player system. Beginning with the psychological motivation for facial behaviour analysis, this field of science has been extensively studied in terms of application and automation. A wide variety of image processing techniques were developed to meet the requirements of the facial expression recognition system. Apart from a theoretical background, this work provides a way to design and implement emotion-based music players. The proposed system will be able to process video of facial behavior, recognize displayed actions in terms of basic emotions, and then play music based on the captured emotions. Major strengths of the system are full automation as well as user and environment independence.

6. FUTURE WORK

In the future, we would like to focus on improving the recognition rate of our system. Also, we would like to develop a mood-enhancing music player in the future that starts with the user's current emotion (which may be sad) and then plays music of positive emotions, thereby eventually giving the user a joyful feeling. The future scope of the system would be to design a mechanism that would be helpful in music therapy treatment, which will help treat patients suffering from disorders like mental stress, anxiety, depression, and trauma. The proposed system also tends to avoid the unpredictable results produced in extreme low-light conditions, and with very poor camera resolution in the future, it will extend to detect more facial features, gestures, and other emotional states (stress level, lie detection, etc.). This project can be used for security purposes in the future. Computers will be able to offer advice in response to the mood of the users. Also, this system will perform various tasks as per the mood of its user. Android development can detect a sleepy mood while driving. Finally, we would like to improve the time efficiency of our system in order to make it more appropriate for use in different applications.

REFERENCES

[1] R. Ramanathan, R. Kumaran, R. Ram Rohan, R. Gupta and V. Prabhu, "An Intelligent Music Player Based on Emotion Recognition," 2017 2nd International Conference on Computational Systems and Information Technology for Sustainable Solution.

[2] S. Deebika, K. A. Indira and Jesline, "A Machine Learning Based Music Player by Detecting Emotions," 2019 Fifth International Conference on Science Technology Engineering and Mathematics (ICONSTEM), Chennai, India, 2019

[3] S. G. Kamble and A. H. Kulkarni, "Facial expression based music player," 2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Jaipur, India, 2016

[4] Darwin C.1998 The expression of the emotions in man and animals, 3rd edn (ed. Ekman P.). London: Harper Collins; New York: Oxford University Press

[5] Ligang Zhang & Dian Tjondronegoro, Dian W. (2011) Facial expression recog-nition using facial movement features. IEEE Transactions on Affective Computing

[6] R. A. Patil, V. Sahula and A. S. Mandal, "Automatic recognition of facial expressions in image sequences: A review," 2010 5th International Conference on Industrial and Information Systems, Mangalore, India, 2010

[7] Xiaojun Wang, Xumin Li, Yong Guan,"Image Edge Detection Algorithm Based on Improved Canny

Operator", Computer Engineering, Vo1.36, No.14, pp. 196-198. Jul. 2012

[8] Menezes, P., Barreto, J.C. and Dias, J. Face tracking based on Haar-like features and Eigenfaces. 5th IFAC Symposium on Intelligent Autonomous Vehicles, Lisbon, Portugal, July 5-7, 2004.

[9] Adolf, F. How-to builds a cascade of boosted classifiers based on Haar-like features. http://robotik.inflomatik.info/other/opencv/OpenCV_ObjectDetection_HowTo.pdf, June 20 2003.

[10] Bradski, G. Computer vision face tracking for use in a perceptual user interface. Intel Technology Journal, 2nd Quarter, 1998.

[11] Lienhart, R. and Maydt, J. An extended set of Haar-like features for rapid object detection. IEEE ICIP 2002, Vol. 1, pp. 900-903, Sep. 2002

[12] Muller, N., Magaia, L. and Herbst B.M. Singular value decomposition, Eigen faces, and 3D reconstructions. SIAM Review, Vol. 46 Issue 3, pp. 518–545. Dec. 2004.

[13] Viola, P. and Jones, M. Rapid object detection using a boosted cascade of simple features. IEEE Conference on Computer Vision and Pattern Recognition, 2001.

[14] G. Chaple and R. D. Daruwala, "Design of Sobel operator based image edge detection algorithm on FPGA," 2014 International Conference on Communication and Signal Processing, Melmaruvathur, India, 2014

[15] The Facial Recognition Technology (FERET) Database. National Institute of Standards and Technology,2003. http://www.itl.nist.gov/iad/humanid/feret

[16] Open Computer Vision Library Reference Manual. Intel Corporation, USA, 2001.