2017 2nd IEEE International Conference On Recent Trends in Electronics Information & Communication Technology (RTEICT), May 19-20, 2017, India

Emotion Based Mood Enhancing Music Recommendation

Aurobind V. Iyer,
Department of Computer Engineering,
K J Somaiya COE,
aurobind.i@somaiya.edu,

Viral Pasad,
Department of Computer Engineering,
K J Somaiya COE,
viral.pasad@somaiya.edu,

Karan Prajapati,
Department of Computer Engineering,
K J Somaiya COE,
karan.prajapati@somaiya.edu and

Smita R. Sankhe,
Department of Computer Engineering,
K J Somaiya COE,
smitasankhe@somaiya.edu

***Abstract*—Music is one of the most effective media as it can instill deep feelings and swamp listeners with subliminal messages. It deftly plays with our emotions which in turn affect our mood. Books, movies and television dramas are a few other media but, in contrast to these, music delivers its message in mere moments. It can aid us when we are feeling low and empower us. When we listen to sad songs, we tend to feel a decline in mood. When we listen to happy songs, we feel happier. Manual classification of songs based on mood, for making of a playlist, is time consuming and labour intensive. Our paper proposes a system 'EmoPlayer', an Android application, which help to minimize these efforts by suggesting the user a list of songs based on his current emotion. The system captures user's image using camera and detects his face. It then detects the emotion and makes a list of songs which will enhance his mood as the songs keep playing. EmoPlayer uses Viola Jones algorithm for face detection and Fisherfaces classifier for emotion classification.**

***Keywords*—Face Detection, Viola Jones, Haar Classifiers, Emotion Detection, Fisherfaces, OpenCV, Computer Vision**

I. INTRODUCTION

Human beings have the natural ability to look at someone's face and guess their mood. This ability if learnt by an electronic device - computer, humanoid robot or a mobile device - can have valuable applications in the real world. Music, a tool for arousing emotions and feelings, is far more powerful than language. Music is something which taps deeply into our emotional core as human beings. Thus, listening to good music can help us elevate our mood from a negative sense to a positive sense. For example, listening to upbeat songs when the person if feeling sad can help him come out of his sadness and start feeling better. This paper proposes one such application, emotion based music recommendation.

Emotion of the user can be easily guessed by looking at his face. For this purpose of face detection and emotion recognition, studying the features from his face is necessary. The problems associated with face detection include background elements, lighting conditions, pose and facial expression. This domain of face detection and emotion detection is currently a very active area of research due to development of Virtual Reality and Augmented Reality. Also, other applications of face detection include focus locking in video recording, monitoring and surveillance systems, humanoid robots and driver alertness system; and application of emotion detection include sentiment analysis system, marketing and feedback evaluation systems. Real-time face detection and recognition systems have limited functionality due to the varying quality of images because of the problems associated like background, illumination, etc. Hence, research and development for solutions related to these problems is an ongoing work.

Using traditional music players, a user had to manually browse through his playlist and select songs that would soothe his mood and emotional experience. This task was labor intensive and an individual often faced the dilemma of landing at an appropriate list of songs. Other systems which detect the mood of the user by using facial expression have their time and memory complexity relatively high and hence fail in achieving a real-time performance. Even if they recognize the mood of the user then their selection of songs for making a playlist is such that it will just pick songs reflecting the current mood of the user and will not try to enhance his mood in any way. So, if the user is sad, he will be provided with a list of songs with sad emotion which can degrade his mood further and can lead to depression. So, the system proposed in this paper will detect the emotion of the user from his facial expressions. It will then provide the user with a playlist of songs, listening to which the user will feel better.

II. LITERATURE SURVEY

Pantic and Rothkrantz [12] proposed system which processes images of frontal and profile face view. Face boundaries have been found using Vertical and horizontal Histogram Analysis. Then, face contour is obtained by thresholding the image with HSV color space values.

Sayali Chavan et al [5] have proposed a system "XBeats-An Emotion Based Music Player" which uses Viola Jones Algorithm for face detection and Support Vector Machine for emotion recognition.

P. Belhumeur et al [1] in their paper "Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection" have proposed a face recognition algorithm, Fisherfaces, based on Fisher's Linear Discriminant. The algorithm has lower error rates as compared to Eigenface.

978-1-5090-3704-9/17/$31.00 © 2017 IEEE

Dolly Reney and Dr.Neeta Tripaathi [7] in their paper "An Efficient Method to Face and Emotion Detection" have detected face from the input image using Viola-Jones face detection algorithm and evaluated the face and emotion detection using KNN classifier.

III. Face Detection

Face detection is an image processing technology that is used in many applications that detects human faces from digital images. Face detection is considered as a special case of object detection. In EmoPlayer, main focus of this technique is on detecting frontal human faces.

A. *Canny Edge Detection*

This technique tries to achieve low error rate, which means a good detection of edges. Canny edge detection includes intricate mathematical calculations, because of which more time is consumed in face detection. EmoPlayer requires fast detection of faces which can be achieved using Viola jones object detection algorithm.

B. *Viola Jones Algorithm*

Viola Jones algorithm is divided in four stages. Firstly, creation of Integral image; Secondly, calculation of the Haar-like features using integral image; Thirdly, The Adaboost algorithm, which is a learning process used to train the classifiers. Finally, we can detect the facial region after cascading multiple strong classifiers.

Step 1: Integral Image

In this step, the captured image will be converted into an integral image. The value of each pixel (x,y) in an integral image is the sum of the pixel values above and to the left of (x,y), inclusive [7].

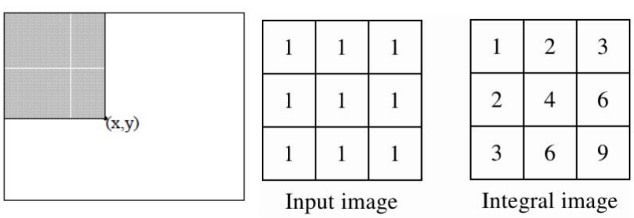

Fig. 1: The Integral Image

$$II(x,y) = \sum_{x' \le x,\, y' \le y} I(x',y'),$$

Where, II(x, y) is the integral image and I(x', y') is the original image.

By doing so, computation of sum of pixels within any rectangle of an image becomes very easy. Sum of pixels can be calculated using just four values of corner of an integral image [7]. This can be demonstrated using following figure. Pixels above and left side of the points B and C has some area in common that is pixels above and left side of point A. Hence while calculating sum of pixels in rectangle ABCD we need to subtract values of B and C from the addition of point A and D.

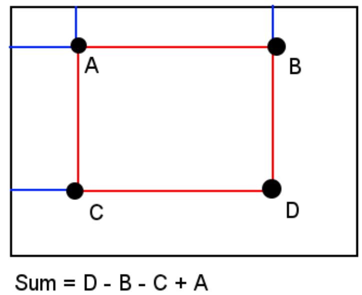

Fig. 2: Sum Calculation

Step 2: Feature Discussion

Similar properties are shared by all human faces; these regularities can be matched using Haar features. Viola Jones algorithm uses these features containing two or more rectangles to analyze the sub-window of the image. To search for the face in the whole image one can move the search window across the image and check every location using the classifier. A single value is obtained from each feature by subtracting the sum of pixels in all white rectangles from the sum of pixels in all black rectangles.

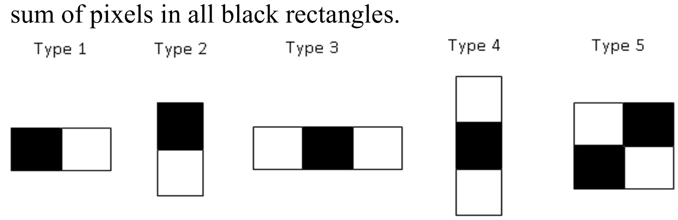

Fig. 3: Haar Features

Step 3: AdaBoost learning Algorithm

In EmoPlayer, variant of AdaBoost is used to select a small set of features and train the classifier. AdaBoost is a machine learning based approach which can construct a strong classifier using a weighted combination of weak classifiers. Output of weak classifier decides whether an image contains a face or not hence the output is binary. This can be determined by the value obtained by each feature. This value is obtained by subtracting sum of pixels in brighter areas from sum of pixels in darker areas.

$$K_i(x) = \begin{cases} 1 & \text{if} \quad p_i f_i(x) < p_i \theta_i \\ 0 & \text{otherwise} \end{cases}$$

where,
$K_i(x)$ – weak classifier,
$f_i(x)$ – calculated value from feature,
p_i – parity: indicates the direction of the inequality sign,
Θ_i – threshold,
x – sub-window of the original image.

Strong classifier is constructed by weighting the multiple weak classifiers. Linear combination of all weak classifiers gives final output for strong classifier.

$$H(x) = sign\left(\sum_{t=1}^{T} \alpha_t h_t(x)\right)$$

where,
a_t is weight applied to classifier 't'.
$h_t(x)$ is output of weak classifier 't'.

Final decision is made by looking at the sign of sum obtained by the given formula.

Step 4: Cascading Classifier:

A single classifier can be adjusted to detect 100% of the faces with a false positive rate of 40%. So, to increase the efficiency of the system, cascade of multiple classifiers can be used. Classifiers in a cascade are arranged in such a way that strength of each classifier in a cascade is more than the previous one. Simpler classifiers are used to reject the majority of sub-windows, after which more complex classifiers are used to achieve low false positive rates. Earlier stage of cascade tries to reject as many negatives as possible. With increase in number of classifiers in a cascade, efficiency of the face detection increases [7].

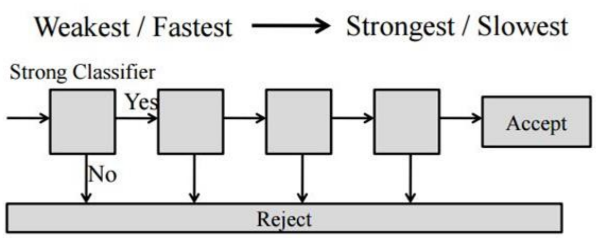

Fig. 4: Cascading classifiers

IV. Emotion Recognition

There are various ways through which human emotion can be recognized such voice, face, body language, etc. We study facial expressions as a means to identify them in this paper. Recognition is the process in which an object is identified into one of the many classes available. Humans express various emotions and we recognize these from their facial images. Images usually lie in high dimensional space which requires large memory for its storage and retrieval and also processing power. Also, the cost of computation on such high dimensional data is often too high [2]. So, we need a technique to reduce the dimensionality of these data. Eigenface and Fisherfaces are methods used for dimensionality reduction.

A. Eigenface method

Eigenface method uses Principal Component Analysis (PCA). It is a statistical technique that is used for data dimension reduction, data compression and also uses eigenvector properties for determining object orientation [4]. It tries to maximize the scatter of all projected samples in the linear projection. However, this maximization is due not only to between-class scatter but also to within-class scatter which results in poor-quality results when used for classification [1].

B. Fisherfaces method

Discriminant Analysis is better approach to reduce dimensionality especially in classification domain. Linear discriminant analysis is one of the most popular Discriminant Analysis. It is a technique used to find a linear combination of features which can be used to separate classes of different objects. It can thus be used for dimensionality reduction as well as a linear classifier.

Fisherfaces method uses fisher's Linear Discriminant (FLD) provides higher between-class scatter as compared to PCA. It focuses to maximize the ratio of between-class to within class scatter. As a result of which the tightly spaced clusters are formed.

As the data set that is used for learning labelled, this information is used to build a more reliable model to reduce dimensionality. It starts by creating an image matrix wherein each column is a vector that represents pixel intensities of the image. Another vector with corresponding class labels is created. The between class and within class scatter is calculated by projecting the image matrix to N-c dimensional subspace, where 'N' is the number of images and 'c' is the number of classes available [2]. Once the scatters have been calculated, LDA is then applied to these scatters.

The method however requires that all the images in the training as well as testing set be of equal size and the faces in these images are in full frontal head position. It works on grey scale vectors of an image rather than working on the image directly. Every image is represented by a weight vector which contains unique attributes necessary for recognition [2]. As described by the authors of [1], Fisherfaces method is insensitive for large variations in lighting conditions. It also has error rate lower than Eigenface technique and the computation time required is also lesser. Hence, we use Fisherfaces model to recognize human emotions.

V. Proposed System

In this paper, we have proposed a system 'EmoPlayer', which is an Android based application. It captures an image of the user using camera of his device and detects the face from this image. The application will then identify the emotion from the face detected. Based on the emotion recognized, it will send the emotion to the music server and will fetch a suitable playlist. The following flowchart (Fig.1) explains it in great detail.

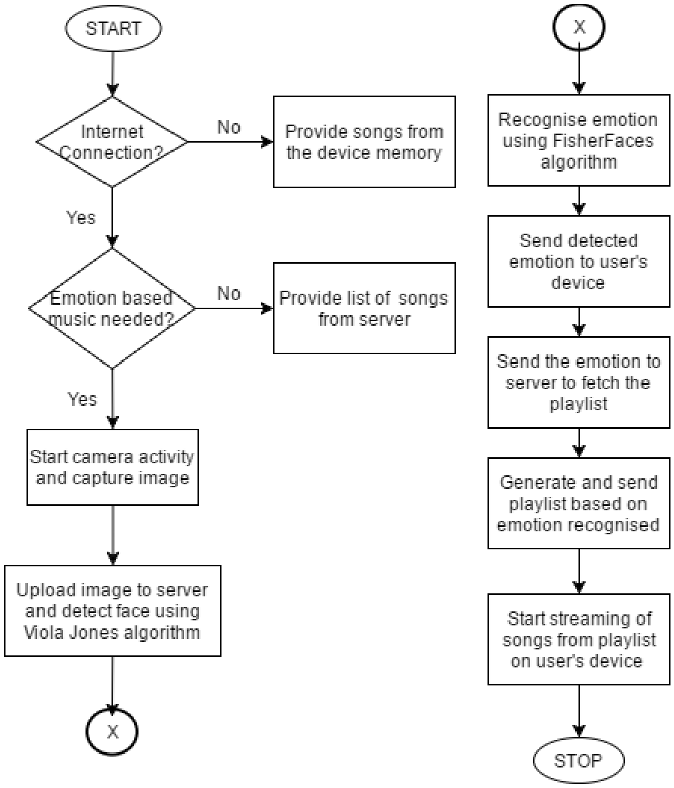

Fig. 5: System workflow

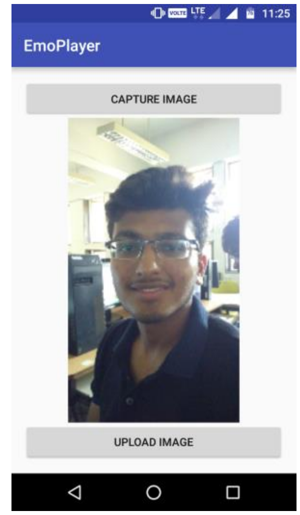

Fig. 6: Capture image

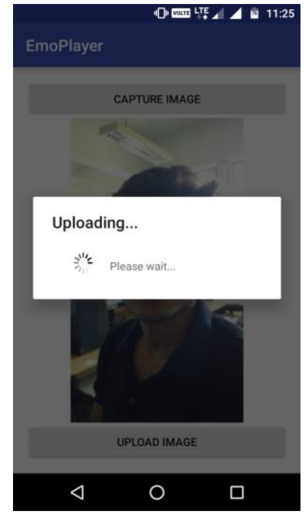

Fig. 7: Upload image

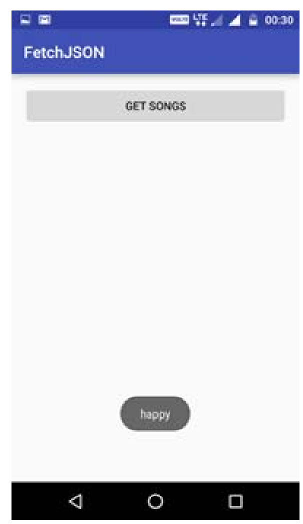

Fig. 8: Display emotion

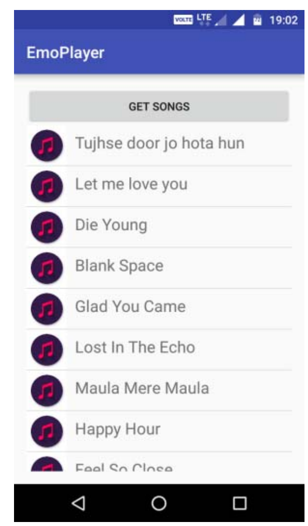

Fig.9: Provide Playlist

As soon as the user launches the application, it checks for internet connectivity. If the user is unable to connect to internet, the application displays the songs from the device's memory. However, on establishment of a stable internet, the user has an option of playing songs either based on his emotion or directly from the music server. Once he opts for an emotion based music recommendation, the app starts the device camera and the user has to capture his image. This image is uploaded to the image server. Face detection is then performed on the server using Viola Jones Algorithm. The code performing this algorithm uses the CascadeClassifier method provided OpenCV library. Next step is emotion recognition which is carried out using Fisherfaces Algorithm. The image server sends the detected emotion to the device, displays it to the user and send emotion to the music server. The music server gets this emotion and uses a code which suggests the user with a list of songs based on this emotion. Songs will be suggested in such an order that the initial few songs will reflect the current emotion and songs down the list will be of a happier emotion. This in turn will influence the mood of the user. Thus, the purpose of the system as a mood enhancing music recommendation is fulfilled.

VI. RESULTS

The simulation has been carried out successfully on multiple Android devices connected to Internet and running on android version 5.0 and above. Attached below are some screenshots of the system from a Moto G 3rd generation smartphone running on android version 6.0.1.

We have used 450 images to train the classifier and test its accuracy. The images used are from CK+ database [8] and some images of our own. The classifier was trained using 80% of the imageset and tested for its accuracy using the rest 20%.

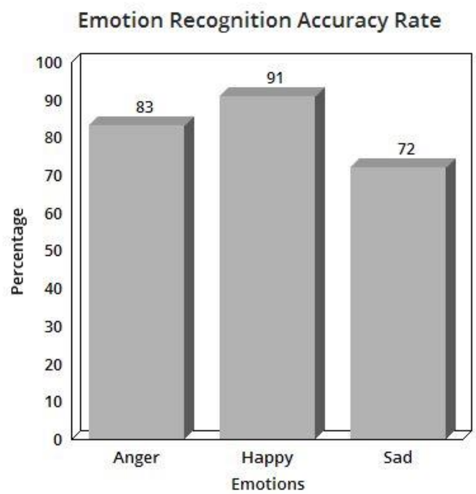

Fig. 10: Emotion Recognition Accuracy rate

CONCLUSION

The point of this paper was to explore the area of facial expression recognition for implementation of an emotion based music player. Manual face analysis utilized by people was immediately supplanted by reasonable computer programming. A wide variety of image processing techniques was developed to meet the facial expression recognition system requirements. Apart from theoretical background, this work gives approaches to outline and execute emotion based music player. Proposed system is able to process the facial image and recognize basic emotions and then play music based on these emotions and also suggest music that enhances the mood of the user. In the future work, we would like to focus on improving the emotion recognition rate of our system and also recognize more different emotions. Also, we'd like to develop an automatic music genre classification system and automatic music emotion classification system [6]. These new modules combined with the existing system will help automate the whole existing system of emotion based mood enhancing music recommendation system.

ACKNOWLEDGMENT

We would like to express our gratitude to our project mentor Mrs. Smita R. Sankhe for her valuable advice and support. We would also like to thank Prof. T. Kanade, Prof. J.F. Cohn and Prof. Y. Tian for providing the Cohn-Kanade database.

REFERENCES

[1] P. Belhumeur, J. Hespanha, and D. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection," IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol. 19, No. 7, pp. 711-720, 1997.

[2] Ashleigh Fratesi, "Automated Real Time Emotion Recognition using Facial Expression Analysis", Master of Computer Science thesis, Carleton University

[3] Mudrov´a, M, Proch´azka, A, "Principal component analysis in image processing", Department of Computing and Control Engineering, Institute of Chemical Technology'.

[4] Paul Viola and Michael J. Jones, "Robust real-time object detection", International Journal of Computer Vision, Vol. 57, No. 2, pp.137–154, 2004.

[5] Sayali Chavan, Ekta Malkan, Dipali Bhatt, Prakash H. Paranjape, "XBeats-An Emotion Based Music Player ", International Journal for Advance Research in Engineering and Technology, Vol. 2, pp. 79-84, 2014.

[6] Xuan Zhu, Yuan-Yuan Shi, Hyoung-Gook Kim and Ki-Wan Eom, "An Integrated Music Recommendation System" IEEE Transactions on Consumer Electronics, Vol. 52, No. 3, pp. 917-925, 2006.

[7] Dolly Reney and Dr.Neeta Tripaathi, "An Efficient Method to Face and Emotion Detection", Fifth International Conference on Communication Systems and Network Technologies, 2015.

[8] Lucey, P., Cohn, J. F., Kanade, T., Saragih, J., Ambadar, Z., & Matthews, I. (2010). The Extended Cohn-Kanade Dataset (CK+): A complete expression dataset for action unit and emotion-specified expression. Proceedings of the Third International Workshop on CVPR for Human Communicative Behavior Analysis (CVPR4HB 2010), San Francisco, USA, 94-101.

[9] Y. Li, S. Wang, Member, IEEE, Y. Zhao, and QiangJi ''Simultaneous Facial Feature Tracking and Facial Expression Recognition'' , Senior Member, IEEE Transactions On Image Processing, Vol. 22, No. 7, July 2013.

[10] M. Pantic and J.M. Rothkrantz, "Facial Action Recognition for Facial Expression Analysis from Static Face Images," IEEE Trans. Systems, Man arid Cybernetics Part B, Vol. 34, No. 3, pp. 1449-1461, 2004.

[11] A. Geetha, V. Ramalingam, S. Palanivel and B. Palaniappan, "Facial Expression Recognition— A Real Time Approach" International Journal of Expert Systems with Applications, Vol. 36, No. 1,2009.

[12] M. Pantic, L. Rothkrantz, "Automatic Analysis of Facial Expressions: The State of the Art", IEEE Transactions On Pattern Analysis and Machine Intelligence, Vol. 22, No. 12, 2000.