

Affective Music Player with FaceSDK for Emotion Recognition

M.J.JeyasheelaRakkini, K.Hemnathan,

School of Computing, SASTRA Deemed University

Abstract—Face is the index of mind and is one of the prominent ways to express one's emotions. Affective computing which can be used to explore the emotional analytics of an individual is used to play the appropriate music based on an individual's emotion is implemented. Face images are retrieved from WEBCAM in the rate of one video frame per 0.5 seconds, the emotions in the face of an individual in these video frames is detected and the music is played in accordance with the emotions in the music player. The facial features are extracted using Luxand FaceSDK which is also used to detect and discern faces from background images. KNN classifier is used for classifying the music based on these emotions. Luxand FaceSDK shows 90% accuracy and the time taken by the KNN classifier to classify the music based on the emotions is 15 seconds.

Keywords—Affective Computing, Face SDK, KNN classification

I. INTRODUCTION

Affective computing, which is also known as Artificial Emotional Intelligence deals with the development of systems that interpret, recognize and analyze human emotions. This Emotional AI is a conglomerate of domains that comprise of computer science, psychology and cognitive science domain. An individual's mood or emotions which are reflected in the facial expressions can be extracted from those expressions and with the extraction of features from eyes, the system can discern between a sleepy person and an awake person. The corresponding music which is relevant to the individual's mood based upon the facial expressions is played in the music player. This work will have promising openings in domains like Human Computer Interaction (HCI), non-intrusive biometric authentication system, touch less logins by looking in the screen, the therapeutic approach in healthcare etc.

II. LITERATURE SURVEY

Srishti Tiwari Dr Aju D et.al describe the facial emotion recognition and how it is exploited to operate or trigger an alert system, after which the individual is

taken care by the concerned personnel in the emergency department[1]. An affective music player which uses deep learning to recommend the music to be played based on the mood of the user is implemented in adaptive music recommendation system which has emotion classifier, music classifier and recommendation module[2]. Emotion music recommendation system which generates a playlist that depends upon the emotion of the user by extracting features from faces, extracting features from audio and mapping the above two modules for recommending the play list[3]. Using Automatic music generation with efficient variable neighborhood search algorithm has been implemented in Morpheus, a framework that produces music for a given tension profile. The affective music is generated by a tension profile and a state of the art pattern detection algorithm is used to detect repeated patterns in a given music template is discussed in [4]. Using the built-in camera in an Android phone, the user image was captured in real time. Grayscale images were used to compress the image files. Cropped features of eye and lip were got and were given to Canny Edge Detection algorithm for edge detection. Emotion recognition was done using Eigen face-based pattern[5]. The content-based features and context-based features of an audio file are extracted and the content based features of an audio file are analyzed for emotions and the mood associated with it[6].

Gabor feature extraction and Neural Network is used to identify emotions in frames got from the real streaming video [7]. AdaBoost algorithm combined with HAAR classifier depth cascade about OpenCV is proposed for face detection algorithm[8]. HAAR Cascades algorithm a machine learning algorithm and an efficient face detection algorithm which is developed by Viola and Jones is used for training a cascade function which numerous positive and negative images[9]. Four new HAAR-like features are used and these features are used with existing HAAR-Like features and given as input to AdaBoost classifier. Thus a powerful classifier for face detection is implemented. This algorithm yields

better face detection results than traditional HAAR-like features algorithm[10].Luxand FaceSDK is a dynamic linking library that can be incorporated and customized in user projects in C#, Visual C++ and Java.The FaceSDK features are leveraged to the optimal level of emotion recognition.

III. WORK FLOW

The face images recognized from a streaming video which is converted to frames in the rate of 0.5 seconds per video frame is given to the preprocessing phase. In the pre-processing phase, grey scale conversion and detection of the face from the background is implemented using FaceSDK. The eye blink detection is done with FaceSDK for checking whether the person is awake or not. The eye aspect ratio of FaceSDK is used for the detection of eyelids whether open or not. The face in the video is localized. The key facial features on the face ROI are detected. The following facial regions like left eye, left eyebrow, right eye, right eyebrow, mouth, nose, jaw are localized and labeled.

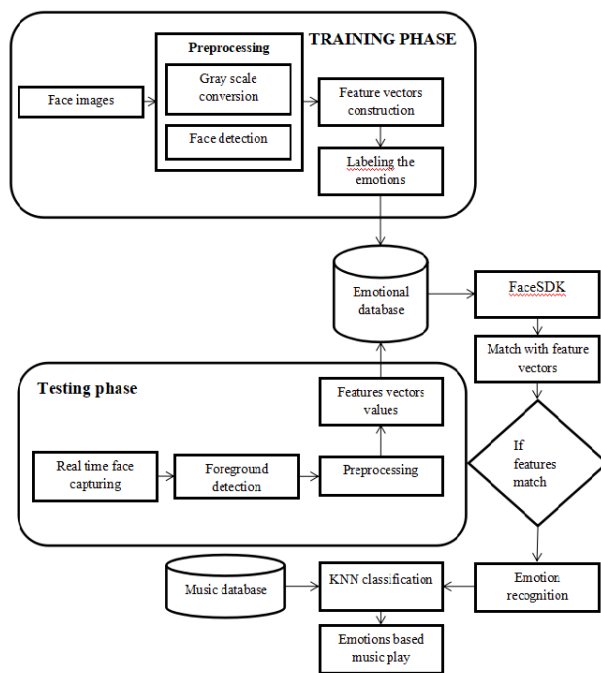


Fig. 1. Flowchart of emotion recognition and affective music player.

The songs are collected and placed in a folder corresponding to each emotion. Using MIR 1.5 Toolbox the rhythm toning feature is extracted and the pitch is

extracted using the Chroma Toolbox. The Auditory Toolbox is used to extract features like 15 MFCC coefficients, centroid, spectral flux, spectral roll-off and kurtosis. The cheerful songs are placed under the happy category. Those songs that resemble depression are placed under sad category. The songs that are quiet and gentle are stored for sleep and are placed under sleep category. The KNN classifier consists of a common weighting scheme. This scheme gives each neighbour a weight of $1/d$, where d is the distance to the neighbour. The neighbours are taken from a set of objects for which the class (for KNN classification) is known. This can be thought of as the training set for the algorithm, though no explicit training step is required. Based on neighborhood values, music is classified under particular emotion and played in the emotional database.

Any new audio song can be classified correctly and placed under the appropriate category by using the KNN classifier. Any new instance y that is the audio song with the above-mentioned features extracted, KNN finds the k neighbours nearest to the instance and places it under appropriate folder for emotions based on Euclidean Distance[11]. $N_k(y)$ represents K nearest neighbors to y and $C(Z)$ represent the class label of Z . Let class $j \in \{1, \dots, \text{class } L\}$. Then

$$N_k^j(y) = \{z \in N_k(y) : c(z) = j\} \quad (1)$$

IV. MODULES DESCRIPTION:

4.1 Facial Image Acquisition:

The face images are captured from real-time streaming video and stored in datasets. The uploaded datasets contain 2D face images. Identification of the faces which are captured by web camera is done in the face identification phase. Here the face image acquisition is done for a particular person and stored.

4.2 Preprocessing:

Preprocessing steps such as grey scale conversion, invert, border analysis, detect edges and region identification is done in the image frames extracted from videos. The Grayscale images, which are also called monochromatic, denoting the presence of only one(mono) colour(chrome). The edge detection is used to analyse the connected curves that indicate the boundaries of objects, the boundaries of surface markings as well as curves that correspond to discontinuities in surface orientation.

4.3 Facial Features extraction:

FaceSDK determines the locations and sizes of human faces in arbitrary(digital) images. It detects facial features and ignores background such as buildings, trees and bodies. Real-time video captured using WEBCAM and frames constructed per five seconds are used to recognize faces. The accuracy rate is high since Face SDK correctly identifies the face and ignores the background images. FSDK_DetectFace function detects the frontal face and stores it in TFacePosition structure. Facial features are detected with FSDK_DetectFacialFeatures()function.

FSDK_DetectEyesinRegion() detects the eyelids whether they are open or not. The Lux and FaceSDK's Dynamic Linking Library (DLL) is used for the recognition of the face. FaceSDK performs well even in poor lighting conditions and returns coordinates of all the faces appearing in a video frame. Here a single individual face is considered for recognition. LBP is not suitable for face recognition from the videos captured so LBP method is not considered here. The disadvantages of LBP method is listed as below [8]

- Not invariant to rotations
- The size of the features increases exponentially with the number of neighbors which leads to an Increase of computational complexity in terms of time and space
- The structural information captured by it is limited.
- Only pixel difference is used, magnitude information ignored.

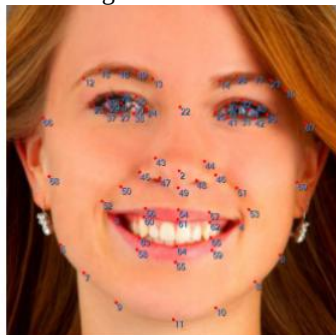


Fig. 2. Facial features of FaceSDK

The seventy features of FaceSDK and its values are got from FaceSDK Dynamic Linking Library (DLL) and used in this affective music player for face recognition[12].

```
public enum FacialFeatures
```

```
{
    FSDKP_LEFT_EYE = 0,
    FSDKP_RIGHT_EYE = 1,
    FSDKP_NOSE_TIP = 2,
    FSDKP_MOUTH_RIGHT_CORNER = 3,
```

```
FSDKP_MOUTH_LEFT_CORNER = 4,
FSDKP_FACE_CONTOUR2 = 5,
FSDKP_FACE_CONTOUR12 = 6,
FSDKP_FACE_CONTOUR1 = 7,
FSDKP_FACE_CONTOUR13 = 8,
FSDKP_CHIN_LEFT = 9,
FSDKP_CHIN_RIGHT = 10,
FSDKP_CHIN_BOTTOM = 11,
FSDKP_LEFT_EYEBROW_OUTER_CORNER = 12,
FSDKP_LEFT_EYEBROW_INNER_CORNER = 13,
FSDKP_RIGHT_EYEBROW_INNER_CORNER = 14,
FSDKP_RIGHT_EYEBROW_OUTER_CORNER = 15,
FSDKP_LEFT_EYEBROW_MIDDLE = 16,
FSDKP_RIGHT_EYEBROW_MIDDLE = 17,
FSDKP_LEFT_EYEBROW_MIDDLE_LEFT = 18,
FSDKP_LEFT_EYEBROW_MIDDLE_RIGHT = 19,
FSDKP_RIGHT_EYEBROW_MIDDLE_LEFT = 20,
FSDKP_RIGHT_EYEBROW_MIDDLE_RIGHT = 21,
FSDKP_NOSE_BRIDGE = 22,
FSDKP_LEFT_EYE_OUTER_CORNER = 23,
FSDKP_LEFT_EYE_INNER_CORNER = 24,
FSDKP_RIGHT_EYE_INNER_CORNER = 25,
FSDKP_RIGHT_EYE_OUTER_CORNER = 26,
FSDKP_LEFT_EYE_LOWER_LINE2 = 27,
FSDKP_LEFT_EYE_UPPER_LINE2 = 28,
FSDKP_LEFT_EYE_LEFT_IRIS_CORNER = 29,
FSDKP_LEFT_EYE_RIGHT_IRIS_CORNER = 30,
FSDKP_RIGHT_EYE_LOWER_LINE2 = 31,
FSDKP_RIGHT_EYE_UPPER_LINE2 = 32,
FSDKP_RIGHT_EYE_LEFT_IRIS_CORNER = 33,
FSDKP_RIGHT_EYE_RIGHT_IRIS_CORNER = 34,
FSDKP_LEFT_EYE_UPPER_LINE1 = 35,
FSDKP_LEFT_EYE_UPPER_LINE3 = 36,
FSDKP_LEFT_EYE_LOWER_LINE3 = 37,
FSDKP_LEFT_EYE_LOWER_LINE1 = 38,
FSDKP_RIGHT_EYE_UPPER_LINE3 = 39,
FSDKP_RIGHT_EYE_UPPER_LINE1 = 40,
FSDKP_RIGHT_EYE_LOWER_LINE1 = 41,
FSDKP_RIGHT_EYE_LOWER_LINE3 = 42
```

```
.....
FSDKP_MOUTH_RIGHT_BOTTOM_INNER = 65,
FSDKP_FACE_CONTOUR14 = 66,
FSDKP_FACE_CONTOUR15 = 67,
FSDKP_FACE_CONTOUR16 = 68,
FSDKP_FACE_CONTOUR17 = 69,}
```

4.4 Emotion classification:

The facial expression recognition is inbuilt in FaceSDK and is used for emotion recognition. The face expression recognition also detects whether the particular individual has his/her eyes open or not and this gives a 100% accuracy finding rate. Five emotions such as anger, happiness with a happy smile, sad smile, sleepy and neutral are recognized. The identified emotions are displayed with emoji. The songs are classified initially and stored in categories like neutral,

sad, happy and sleep (Fig.2 & Fig.3).The age and gender can be shown by FaceSDK accurately.

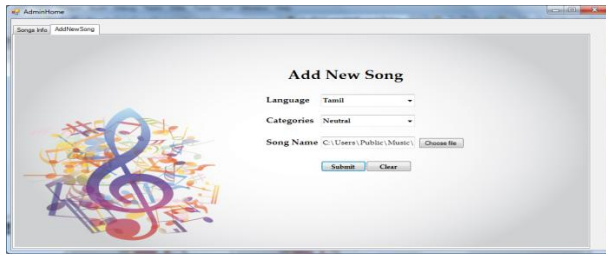


Fig. 3. Addition of songs to the MS SQL Server Database

Fifteen frames are analyzed for valid emotion recognition and after that, the audio song collection based on the emotion is played. An appropriate message such as a cheer up message in case of sad emotion detection can be played before the appropriate collection of audio songs are played one after the other. The user need not press the start camera button or the play music button explicitly. The events take place automatically while the code gets executed.

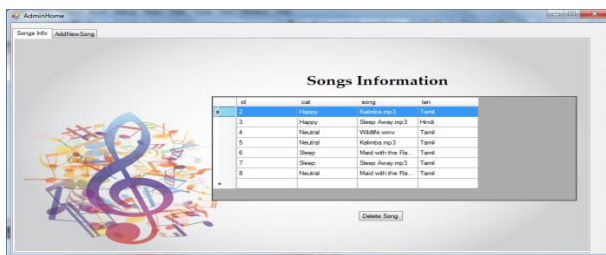


Fig. 4. Categorization of songs

The collection of songs are placed in Microsoft SQL server 8.0 and the songs which are in the category are played one after the other according to the time interval set by user or till the user presses the stop button. Mood taxonomy is used to distinguish the moods of a person whether he/she is happy or sad.



Fig. 5. Recognition of Sad emotion

The WEBCAM starts to capture video and after twenty-five seconds, the emotion is recognized. The face recognized in the frame which has eyes open as 99% and sad smile which is 7% (according to the lip coordinates) is classified under sad emotion and the song which has been grouped under the category sad is played accordingly (Fig 4 & 5).

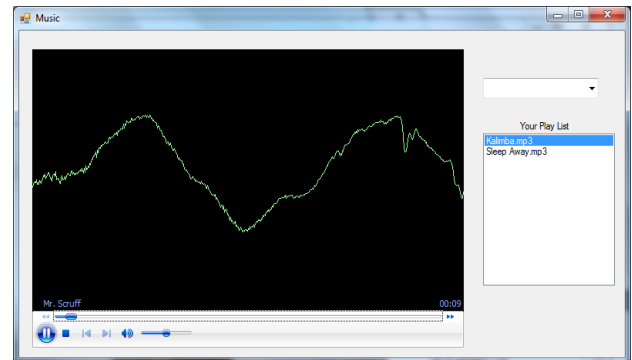


Fig. 6. The audio song played for sad emotion

Sad emotion is recognized accurately and the audio song is played appropriately. The facial expression recognition in FaceSDK helps recognize the facial emotions. Happy smile is recognized when the lip features vary from 60% to 100% and the eyelid feature extracted shows 100% (the get value confidence). DirectX compatible webcams that work in Windows are supported by this Luxand FaceSDK. MJPEG IP cameras, AXIS cameras are also supported by this FaceSDK. This makes the application suitable for security, surveillance and face based authentication. With IP cameras, face images and expressions are got from remote cameras. The user login in web cam can be automated.

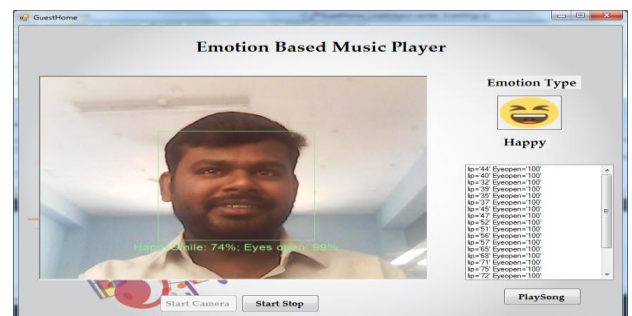


Fig. 7. Happy emotion recognition

A happy emoji which is a PNG (Portable Network Graphics) file can be displayed in picture box when the happy emotion is detected. The happy song is played after that particular emotion is detected accurately. It

takes a time period of 25 seconds and analysis of 50 frames.

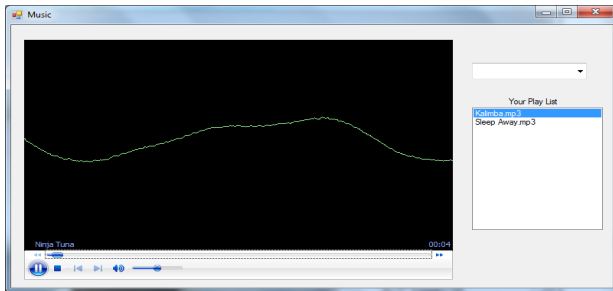


Fig. 8. Audio song played for Happy emotion

The appropriate audio song is played from the particular category of songs after the happy emotion is recognized.



Fig. 9. The Eyelid recognition for sleep.

The eyelid feature which has less than 20% is classified under “sleep” and the emoji is displayed appropriately.

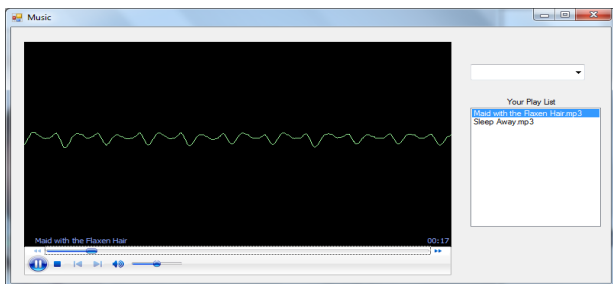


Fig . 10. The appropriate song is played for recognition of sleep with eyelids closed for a user given time interval of 25 seconds.

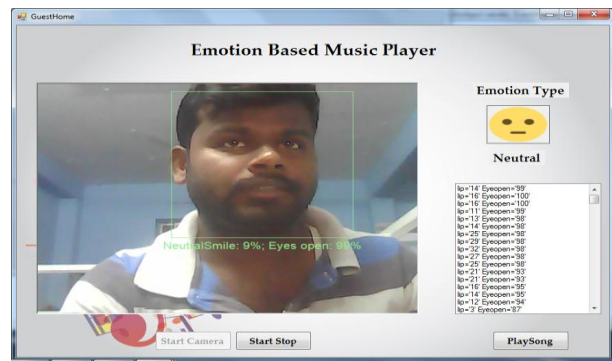


Fig. 11. Neutral emotion recognition

The neutral emotion and neutral smile is recognized with 99% for eyelids open and 9% for smile recognized from lips is classified under Neutral emotion. The emoji is displayed appropriately. The templates extracted from faces can be used stored in database and the function `FSDK_matchfaces()` function is used to match faces. If the similarity level is greater than 0.99%, the probability of showing the correct person is high[13].

The new augmented reality given by Luxand FaceSDK augments the reality got by extracting 66 facial points from faces recognized from video stream. The mirror reality SDK can make the individual's recognized face appear fat, look like aged person, an anorexic, a zombie, a baby, with a makeup look and without a makeup look. Thus an ordinary webcam is turned into a magic mirror, where this magic reality with amazing transformations can be viewed. This has many applications in entertainment industry, webmasters of social networks and game developers.

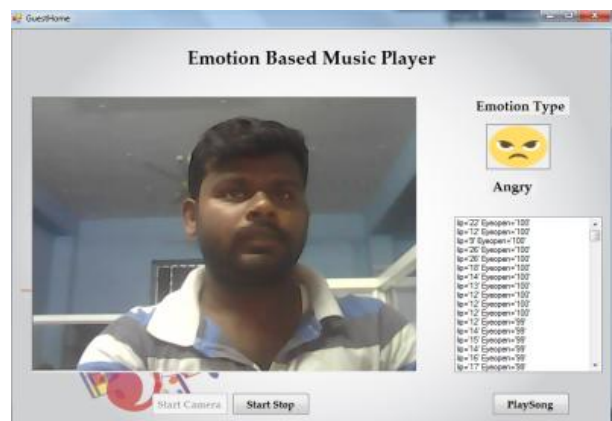


Fig. 12. Anger emotion recognition

The anger emotion is recognised with 100% for open eyelids and the confidence value of lips is 17. Then the

angry emoji is displayed appropriately. The accuracy histogram graph of FaceSDK is more accurate than the existing classifiers.

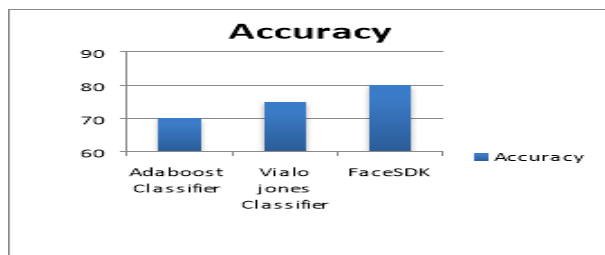


Fig. 13. Accuracy graph.

The Histogram graph of FaceSDK for emotion recognition for all emotions has more accuracy than the conventional classifiers for emotions like Viola Jones Classifier and AdaBoost Classifier. The false rejection rate is 6.1% in still images and the false acceptance rate is 0.1% based on the face recognition grand challenge test. The gender recognition is 93% in still pictures and 97% in motion streams.

TABLE 2. Accuracy table for Emotion classifiers.

Classifier name	Accuracy
AdaBoost Classifier	70%
Vialo Jones Classifier	75%
FaceSDK	80%

The FaceSDK classifier is more accurate in identifying the faces and classifying the emotions.

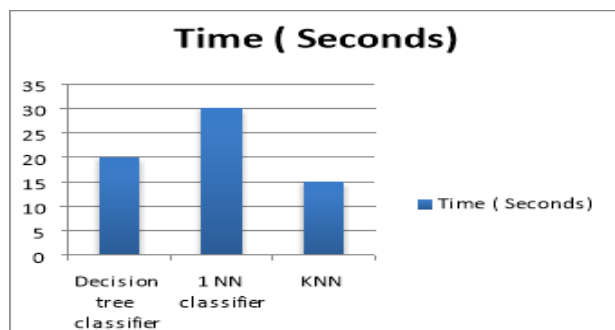


Fig.14. KNN classifier for Audio Songs.

The KNN classifier which is used for audio songs classification and retrieval from database according to a particular category of emotions takes much less time than the other classifiers.

TABLE 1: Automation of playlist with classifiers

Classifier name	Time (Seconds)
Decision tree classifier	20
1 NN classifier	30
KNN	15

KNN algorithm is used to classify the music based on emotions classified by previous modules. Based on neighbourhood values, the songs are classified and played in the music player. The playlist is automated by KNN classifier [14][15].

V. Conclusion

Luxand FaceSDK is a powerful face detection and face feature recognition dynamic linking library that can be linked with Visual Studio C#, Java applications and can be used for emotion analysis and recognition. This cross-platform FaceSDK can be integrated into any application with ease. Direct show compatible USB cameras are supported by this FaceSDK. The age and gender of an individual can be displayed in this application facilitated by FaceSDK. The five different emotions are classified with FaceSDK more accurately than the other emotions. Recognition of multiple faces in a particular image frame from live streaming videos and their emotions can be implemented as a future work with FaceSDK.

References

- [1] "Operating an Alert System using Facial Expression", Srishti Tiwari, Dr. Aju. D, International Conference on Innovations in Power and Advanced Computing Technologies [i-PACT2017].
- [2] "Smart Music Player Integrating Facial Emotion Recognition and Music Mood Recommendation", Shlok Gilda, Husain Zafar, Chintan Soni and Kshitija Waghurdekar, International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), 2017.
- [3] "Emotion Based Music Player", Hafeez Kabani, Sharik Khan, Omar Khan, Shabana Tadv, International Journal of Engineering Research and General Science Volume 3, Issue 1, January-February, 2015.
- [4] "Morpheus: generating structured music with constrained patterns and tension", Dorian Herremans, Member, IEEE, and Elaine Chew, Member, IEEE transactions on affective computing vol 8, no 3, july-september 2017.
- [5] "Real Time Emotion Based Music Player for Android", D.M.M.T. Dissanayaka¹, S.R. Liyanage², Proceedings of International Postgraduate Research Conference 2015, University of Kelaniya

- [6] "EMOTION DETECTION OF AUDIO FILES", 3RD INTERNATIONAL CONFERENCE ON COMPUTING FOR SUSTAINABLE GLOBAL DEVELOPMENT (INDIA COM), 2016.
- [7] "FACIAL EXPRESSION RECOGNITION USING GABOR WAVE", MAHESH KUMBHAR, MANASI PATIL, ASHISH JADHAV, INTERNATIONAL JOURNAL OF COMPUTER APPLICATIONS (0975 – 8887) VOLUME 68– NO.23, APRIL 2013.
- [8] "A scale- and orientation-adaptive extension of Local Binary Patterns for texture classification", Hegenbart.S, Uhl.A, ScienceDirect, Volume 48, Issue 8, August 2015, Pages 2633-2644.
- [9] "Automatic Emotion Detection Model from FacialExpression", Debishree Dagar, Abir Hudait, H. K. Tripathy, M. N. Das, 2016 International Conference on Advanced Communication Control and Computing Technologies (ICACCCT)
- [10] "A depth cascade face detection algorithm based on ADABOOST", WenXiang Yu, Jiapeng Jiu, Chen Liu, ZhengQiu Yang, IEEE International Conference on Network Infrastructure and Digital Content (IC-NIDC), 2016.
- [11] "Audio-Visual Speech Recognition with Weighted KNN based Classification in Mandarin Database", Tsang-Long Pao¹, en Yuan Liao², Yu-Te Chen³. Third International Conference on Intelligent Information Hiding and Multimedia Signal Processing, 2007. IHHMSP 2007.
- [12] <https://www.luxand.com/facesdk>
- [13] Sekar, K.R., Ravichandran, K.S., Krishankumar, R., "Host environment focused software component quality ranking and selection", (2015) Global Journal of Pure and Applied Mathematics, 11 (6), pp. 3991-4004.
- [14] Sekar, K.R., Ravichandran, K.S., Krishankumar, R., "Multi service software components selection based on APRIORI and similarity measures", (2015) Global Journal of Pure and Applied Mathematics, 11 (5), pp. 3777-3791.
- [15] Sekar, K.R., Ravichandran, K.S., Sethuraman, J., Jangiti, S., "DMK-medoid heuristic product ranking in online market", (2014) International Journal of Applied Engineering Research", 9 (19), pp. 5895-5912.

