International Journal of Engineering Research and General Science Volume 3, Issue 1, January-February, 2015
ISSN 2091-2730

Emotion Based Music Player

Hafeez Kabani[1], Sharik Khan[2], Omar Khan[3], Shabana Tadvi[4]

[1]Department of Computer Science and Engineering

[2]Department of Computer Science and Engineering

[3]Department of Computer Science and Engineering

[4]Asst. Professor, Department of Computer Science and Engineering

M.H Saboo Siddik College of Engineering, University of Mumbai, India

Email:-kabani152@gmail.com

Abstract— The human face is an important organ of an individual's body and it especially plays an important role in extraction of an individual's behavior and emotional state. Manually segregating the list of songs and generating an appropriate playlist based on an individual's emotional features is a very tedious, time consuming, labor intensive and upheld task. Various algorithms have been proposed and developed for automating the playlist generation process. However the proposed existing algorithms in use are computationally slow, less accurate and sometimes even require use of additional hardware like EEG or sensors. This proposed system based on facial expression extracted will generate a playlist automatically thereby reducing the effort and time involved in rendering the process manually. Thus the proposed system tends to reduce the computational time involved in obtaining the results and the overall cost of the designed system, thereby increasing the overall accuracy of the system. Testing of the system is done on both user dependent (dynamic) and user independent (static) dataset. Facial expressions are captured using an inbuilt camera. The accuracy of the emotion detection algorithm used in the system for real time images is around 85-90%, while for static images it is around 98-100%.The proposed algorithm on an average calculated estimation takes around 0.95-1.05 sec to generate an emotion based music playlist. Thus, it yields better accuracy in terms of performance and computational time and reduces the designing cost, compared to the algorithms used in the literature survey.

Keywords— Audio Emotion Recognition, Music Information Retrieval, Emotion Extraction Module, Audio Feature Extraction Module, Artificial Neural Networks, Confusion Matrix, Viola and Jones Face Detection.

I. INTRODUCTION

Music plays a very important role in enhancing an individual's life as it is an important medium of entertainment for music lovers and listeners and sometimes even imparts a therapeutic approach. In today's world, with ever increasing advancements in the field of multimedia and technology, various music players have been developed with features like fast forward, reverse, variable playback speed (seek & time compression),local playback, streaming playback with multicast streams. Although these features satisfy the user's basic requirements, yet the user has to face the task of manually browsing through the playlist of songs and select songs based on his current mood and behaviour. The introduction of Audio Emotion Recognition (AER) and Music Information Retrieval (MIR) in the traditional music players provided automatically parsing the playlist based on various classes of emotions and moods.

AER is a technique which deals with classifying a received audio signal, by considering its various audio features into various classes of emotions and moods, whereas MIR is a field that extracts some critical information from an audio signal by exploring some audio features like pitch ,energy,MFCC, flux etc.Though both AER and MIR included the capabilities of avoiding manual segregation of songs and generation of playlist, yet it is unable to incorporate fully a human emotion controlled music player. Although human speech and gesture are a common way of expressing emotions, but facial expression is the most ancient and natural way of expressing feelings, emotions and mood.

The main objective of this paper is to design an efficient and accurate algorithm that would generate a playlist based on current emotional state and behaviour of the user. The algorithm designed requires less memory overheads, less computational and processing time, reducing the cost of any additional hardware like EEG or sensors. The facial expression would categorize into 5 different types

of facial expressions like anger, joy, surprise, sad, and excitement. A high accurate audio extraction technique is proposed that extracts significant, critical and relevant information from an audio signal based on certain audio features in a much lesser time. An emotion model is proposed that classifies a song based on any of the 7 classes of emotions viz sad, joy-anger, joy-surprise, joy-excitement, joy, anger, and sad-anger. The emotion extraction module and audio feature extraction module is combined using an Emotion-Audio integration module. The proposed mechanism achieves a better efficiency and real time performance than the existing methodologies.

This paper is organized into: Section 2 gives the brief study of literature survey. Section 3 explains the methodology; Section 4 provides the experimental analysis and results. Section 5 gives the conclusion of the paper and future work.

II.LITERATURE SURVEY

I .Various techniques and approaches have been proposed and developed to classify human emotional state of behavior. The proposed approaches have focused only on the some of the basic emotions. For the purpose of feature recognition, facial features have been categorized into two major categories such as Appearance-based feature extraction and Geometric based feature extraction by zheng et. al [17]. Geometric based feature extraction technique considered only the shape or major prominent points of some important facial features such as mouth and eyes. In the system proposed by Changbo et. al [2], around a total of 58 major landmark points was considered in crafting an ASM..The appearance based extraction feature like texture, have also been considered in different areas of work and development. An efficient method for coding and implementing extracted facial features together with multi-orientation and multi-resolution set of Gabor filters was proposed by Michael Lyons [10] et. al.

II. An accurate and efficient statistical based approach for analyzing extracted facial expression features was proposed by Renuka R. Londhe et al. [13]. The paper was majorly focused on the study of the changes in curvatures on the face and intensities of corresponding pixels of images. Artificial Neural Networks (ANN) was used in the classification extracted features into 6 major universal emotions like anger, disgust, fear, happy, sad, and surprise. A Scaled Conjugate Gradient back-propagation algorithm in correlation with two-layered feed forward neural network was used and was successful in obtaining a 92.2 % recognition rate. In order to reduce the human effort and time needed for manual segregation of songs from a playlist, in correlation with different classes of emotions and moods, various approaches have been proposed.

III. Thayer [16] proposed a very useful 2-dimenesional (Stress v/s energy) model plotted on two axes with emotions depicted by a 2-dimensional co-ordinate system, lying on either 2 axes or the 4 quadrants formed by the 2-dimensional plot. The music mood tags and A-V values from a total 20 subjects were tested and analyzed in Jung Hyun Kim's[7] work, and based on the results obtained from the analysis, the A-V plane was classified into 8 regions(clusters), depicting mood by data mining efficient k-means clustering algorithm.

IV. Numerous approaches have been designed to extract facial features and audio features from an audio signal and very few of the systems designed have the capability to generate an emotion based music playlist using human emotions and the existing designs of the systems are capable to generate an automated playlist using an additional hardware like Sensors or EEG systems thereby increasing the cost of the design proposed. Some of the drawbacks of the existing system are as follows

i. Existing systems are very complex in terms of time and memory requirements for extracting facial features in real time.
ii. Based on the current emotional state and behavior of a user, existing systems possess a lesser accuracy in generation of a playlist.
iii. Some existing systems tend to employ the use of human speech or sometimes even the use of additional hardware for generation of an automated playlist, thereby increasing the total cost incurred.

This paper primarily aims and focuses on resolving the drawbacks involved in the existing system by designing an automated emotion based music player for the generation of customized playlist based on user extracted facial features and thus avoiding the employment of any additional hardware. It also includes a mood randomized and appetizer function that shifts the mood generated playlist to another same level of randomized mood generated playlist after some duration.

III. METHODOLOGY

The proposed algorithm in this involves an emotion music recommendation system that provides the generation of a customized playlist in accordance to the user's emotional state. The proposed system involves three major modules: Emotion extraction module, Audio feature extraction module and an Emotion-Audio recognition module. Emotion extraction module and Audio feature extraction module are two separate modules and Emotion-Audio recognition module performs the mapping of modules by querying the audio meta-data file. Fig 1 illustrates block diagram of proposed system.

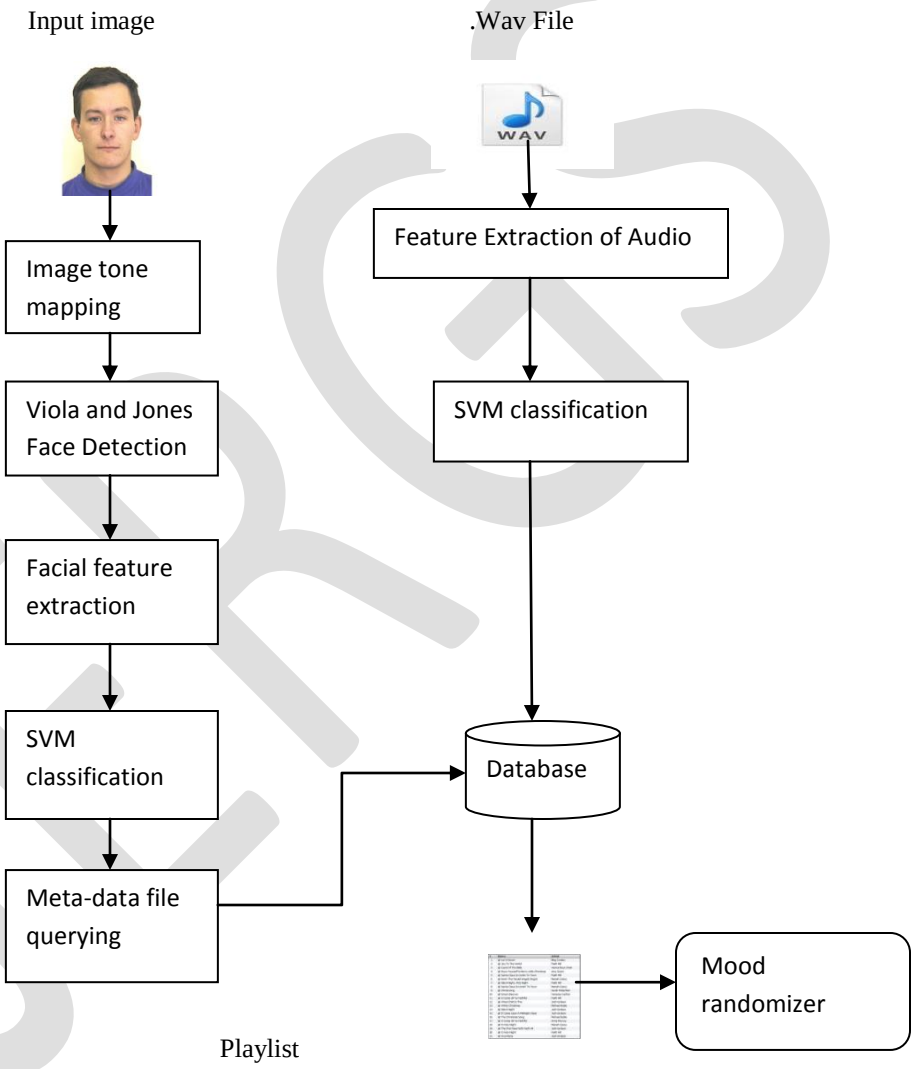

Fig 1 Block Diagram of Proposed Algorithm

A.EMOTION EXTRACTION MODULE:

Image of a user is captured using a webcam or it can be accessed from the stored image in the hard disk. This acquired image undergoes image enhancement in the form of tone mapping in order to restore the original contrast of the image.After image enhancement all images are converted into binary image format and the face is detected using Viola and Jones algorithm where the 'Frontal Cart property' of the algorithm is used that only detects upright and face forwarding features with a maximum threshold value set in the range of 16-20. The output of Viola and Jones Face detection block forms an input to the facial feature extraction block.

To increase the accuracy and an aim to obtain real time performance only features of eyes and mouth are appropriate enough to depict the emotions accurately. For extracting the features of mouth and eyes certain calculations and measurements are taken into consideration. Equations (1), (2), (3) and (4) illustrate the bounding box calculations for extracting features of a mouth.

X (start pt of mouth) =X (mid pt of nose) – (X (end pt of nose) – (Xstart pt of nose)) (1)

X (end pt of mouth) =X (mid pt of nose) + ((Xend pt of nose) – (Xstart pt of nose)) (2)

Y (start pt of mouth) =Y (mid pt of nose) +15 (3)

Y (end pt of mouth) =Y (start pt of mouth) +103 (4)

Where (X (start pt of mouth), Y (start pt of mouth)) and (X (end pt of mouth), Y (end pt of mouth)) illustrates start and end points of the bounding box for mouth respectively, (X (mid pt of nose), Y (mid pt of nose)) illustrates midpoint of noise and ((Xend pt of nose), (Xstart pt of nose)) illustrates end and start point of noise. Classification is performed using Support Vector Machine (SVM) which classifies it into 7 classes of emotions.

B.AUDIO FEATURE EXTRACTION MODULE:

In this module a list of songs forms the input. As songs are audio files, they require a certain amount of preprocessing Stereo signals obtained from the Internet are converted to 16 bit PCM mono signal around a variable sampling rate of 48.6 kHz. The conversion process is done using Audacity technique.

The pre-processed signal obtained undergoes an audio feature extraction, where features like rhythm toning is extracted using MIR 1.5 Toolbox, pitch is extracted using Chroma Toolbox and other features like centroid, spectral flux, spectral roll off, kurtosis,15 MFCC coefficients are extracted using Auditory Toolbox.

Audio signals are categorized into 8 types viz. sad, joy-anger, joy-surprise, joy-excitement, joy, anger, sad-anger and others.

1. Songs that resemble cheerfulness, energetic and playfulness are classified under joy.
2. Songs that resemble very depressing are classified under the sad.
3. Songs that reflect mere attitude, revenge are classified under anger.
4. Songs with anger in playful is classified under Joy-anger category.
5. Songs with very depress mode and anger mood are classified under Sad-Anger category.
6. Songs which reflect excitement of joy is classified under Joy-Excitement category.
7. Songs which reflect surprise of joy is classified under Joy-surprise category.
8. All other songs fall under ‘others’ category.

C.EMOTION-AUDIO INTEGRATION MODULE:

Emotions extracted for the songs are stored as a meta-data in the database. Mapping is performed by querying the meta-data database. The emotion extraction module and audio feature extraction module is finally mapped and combined using an Emotion-Audio integration module. Fig 2 illustrates mapping of Facial features and Audio features. For example, if an input facial image is categorized under joy, the system will display songs under joy, joy-anger, Joy-Excitement, Joy-surprise category.

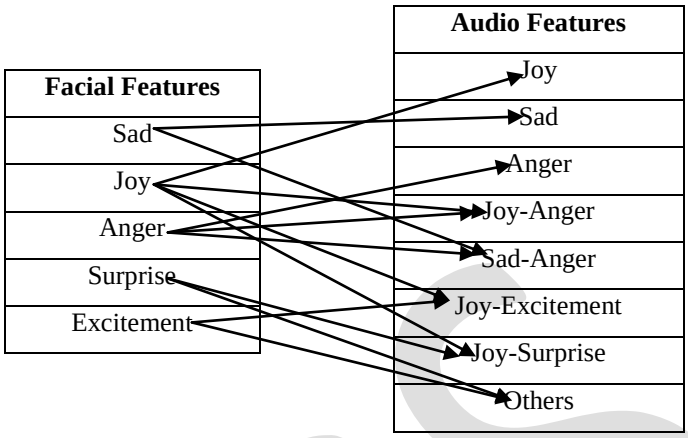

Fig 2 Modules mapping

IV.RESULTS AND EXPERIMENTS:

Testing and implementation is performed using either MATLAB R2013a or latest 2014 version of MATLAB on Windows7/8, 32 bit operating system and Intel i3 core processor. Facial expression extraction is done on both user independent and dependent dataset. A dataset consisting of facial image of 25 individuals was selected for user independent experiment and dataset of 10 individuals was selected for user dependent experimentation. An image of size 4000X3000 was used for static and dynamic dataset experiment.

A.EMOTION EXTRACTION:

A user independent dataset of 30 images and user dependent dataset of 5 images is selected for extraction of emotions. Estimated time for various modules of Emotion Extraction Module is illustrated in Table 1.

Table 1 Time Estimation of Various modules of Emotion Extraction Module

Module	Time Taken(sec)
Face Detection	0.8126
Facial Feature Extraction	0.9216
Classification using SVM	0.1956
Emotions	0.9994

B.AUDIO FEATURE EXTRACTION:

A dataset of around 200 songs was considered for experimentation and testing of audio feature extraction module and the songs were collected from various Bollywood music sites like Djmaza.in, Songs.pk etc. Estimated accuracy for various categories of emotions is depicted in Table 2.

Table 2 Estimated Accuracy for different categories of Audio Feature

Emotions	Accuracy
Joy	89%
Sad	99%
Anger	99.8%
Surprise	88%
Excitement	95%
Joy- Excitement	96.4%
Joy-Surprise	100%

C.EMOTION BASED MUSIC PLAYER:

The Proposed system is tested and experimented against an in-built camera, thus the total cost involved in implementation is almost negligible. Average estimated time for various modules of propped system is illustrated in Table 3.

Table 3 Average Time Estimation of Various modules of Proposed System

Module	Time Taken(sec)
Emotion Extraction Module	0.9994
Emotion-Audio Integration Module	0.0006
Proposed System	1.0000

IV. CONCLUSION AND FUTURE SCOPE:

Experimental results have shown that the time required for audio feature extraction is negligible (around 0.0006 sec) and songs are stored pre-handed the total estimation time of the proposed system is proportional to the time required for extraction of facial features (around 0.9994 sec).Also the various classes of emotion yield a better accuracy rate as compared to previous existing systems. The computational time taken is 1.000sec which is very less thus helping in achieving a better real time performance and efficiency.

The system thus aims at providing the Windows operating system users with a cheaper, additional hardware free and accurate emotion based music system. The Emotion Based Music System will be of great advantage to users looking for music based on their mood and emotional behavior. It will help reduce the searching time for music thereby reducing the unnecessary computational time and thereby increasing the overall accuracy and efficiency of the system. The system will not only reduce physical stress but will also act as a boon for the music therapy systems and may also assist the music therapist to therapize a patient. Also with its additional features mentioned above, it will be a complete system for music lovers and listeners.

The future scope in the system would to design a mechanism that would be helpful in music therapy treatment and provide the music therapist the help needed to treat the patients suffering from disorders like mental stress, anxiety, acute depression and trauma.
The proposed system also tends to avoid in future the unpredictable results produced in extreme bad light conditions and very poor camera resolution.

REFERENCES:

[1]. Chang, C. Hu, R. Feris, and M. Turk, “Manifold based analysis of facial expression,” Image Vision Comput ,IEEE Trans. Pattern Anal. Mach. Intell. vol. 24, pp. 05–614, June 2006.

[2]. A. habibzad, ninavin, Mir kamalMirnia," A new algorithm to classify face emotions through eye and lip feature by using particle swarm optimization."

[3]. Byeong-jun Han, Seungmin Rho, Roger B. Dannenberg and Eenjun Hwang, "SMERS: music emotion recognition using support vector regression", 10thISMIR , 2009.

[4]. Alvin I. Goldmana, b.Chandra and SekharSripadab, "Simulationist models of face-based emotion recognition".

[5]. Carlos A. Cervantes and Kai-Tai Song , "Embedded Design of an Emotion-Aware Music Player", IEEE International Conference on Systems, Man, and Cybernetics, pp 2528-2533 ,2013.

[6]. Fatma Guney, "Emotion Recognition using Face Images", Bogazici University, Istanbul, Turkey 34342.

[7]. Jia-Jun Wong, Siu-Yeung Cho," Facial emotion recognition by adaptive processing of tree structures".

[8]. K.Hevener,"The affective Character of the major and minor modes in music", The American Journal of Psychology,Vol 47(1) pp 103-118,1935.

[9]. Samuel Strupp, Norbert Schmitz, and KarstenBerns, "Visual-Based Emotion Detection for Natural Man-Machine Interaction".

[10]. Michael lyon and Shigeru Akamatsu, "Coding Facial expression with Gabor wavelets.", IEEE conf. on Automatic face and gesture recognition, March 2000

[11]. Kuan-Chieh Huang, Yau-Hwang Kuo, Mong-Fong Horng ,"Emotion Recognition by a novel triangular facial feature extraction method".

[12]. S. Dornbush, K. Fisher, K. McKay, A. Prikhodko and Z. Segall "Xpod- A Human Activity and Emotion Aware Mobile Music Player", UMBC Ebiquity, November 2005.

[13]. Renuka R. Londhe, Dr. Vrushshen P. Pawar, "Analysis of Facial Expression and Recognition Based On Statistical Approach", International Journal of Soft Computing and Engineering (IJSCE) Volume-2, May 2012.

[14]. Russell, "A circumplex model of affect", Journal of Personality and Social Psychology Vol-39(6), pp1161–1178 , 1980.

[15]. Sanghoon Jun, Seungmin Rho, Byeong-jun Han and Eenjun Hwang, "A fuzzy inference-based music emotion recognition system",VIE,2008.

[16]. Thayer " The biopsychology of mood & arousal", Oxford University Press ,1989.

[17]. Z. Zeng "A survey of affect recognition methods: Audio, visual, and spontaneous expressions,"IEEE. Transaction Pattern Analysis, vol 31, January 2009