MoodyPlayer: A Mood based Music Player

Abhishek R. Patel
NBN Sinhgad School of Engineering
Pune, Maharashtra,
India

Anusha Vollal
NBN Sinhgad School of Engineering
Pune, Maharashtra,
India

Pradnyesh B. Kadam
NBN Sinhgad School of Engineering
Pune, Maharashtra,
India

Shikha Yadav
NBN Sinhgad School of Engineering
Pune, Maharashtra,
India

Rahul M. Samant
NBN Sinhgad School of Engineering
Pune, Maharashtra,
India

ABSTRACT
Increasing and maintaining human productivity of different tasks in stressful environment is a challenge. Music is a vital mood controller and helps in improving the mood and state of the person which in turn will act as a catalyst to increase productivity. Continuous music play requires creating and managing personalized song playlist which is a time consuming task. It would be very helpful if the music player itself selects a song according to the current mood of the user. The mood of the user can be detected by a facial expression of the person. A facial expression detection system should address three major problems: detection of face from an image, facial feature extraction and facial expression classification[1].The first stage is of face detection from an image for which various techniques used are model based face tracking which includes real-time face detection using edge orientation matching [2], Robust face detection using Hausdorff distance [3], weak classifier cascade which includes Viola and Jones algorithm [4], and Histograms of Oriented Gradients (HOG) descriptors. The next stage is to extract features from detected face. Two major approaches for feature extraction which use Gabor filters [Dennis Gabor] and Principle Component Analysis [Jolliffe]. The final stage is of image classification for mood detection, where various classifiers like BrownBoost [Freund, 2001], AdaBoost [Freund and Schapire, 1995] and Support Vector Machines (SVM) are available. The proposed system will use classic Histograms of Oriented Gradients (HOG) along with facial landmark detection technique; these detected features then passed through SVM classifier to predict the mood of the user. This predicted mood will stimulate the creation of playlist.

General Terms
Pattern Recognition, Image Classification, Pattern Matching, Emotion Recognition,

Keywords
Image Processing, Machine Learning, Face Detection, Mood Detection, Facial Expression Recognition, Histograms of Oriented Gradients (HOG), Support Vector Machine.

1. INTRODUCTION
This project is based on the principle of detection of human emotions using image processing, and to play music which is appropriate for enhancing that emotional state. It works when mathematical operations are performed using the framework of signal processing which uses an image or a series of images as input.

The state of mind and current emotional mood of human beings can be easily observed through their facial expressions. The different muscles beneath the face act as action units which signal different emotions. The Institute of Neuroscience and Psychology researched that 'wide open eyes' were the signals of happy and sad nature, whereas the 'wrinkled nose' depicted anger or disgust. This project was made by taking three basic emotions (happy, sad, and neutral) into consideration [5].

The face detection in this project is made by creation of face chips on the dimensions of a human face by creating and joining multiple feature points on chin, cheeks, lips, forehead, etc. An algorithm is created and inserted into the system which classifies the emotions into their respective order based on various machine learning techniques.

Music is often described as a 'language of emotions' throughout the globe. Let it be a 80 year old man or a 12 year old girl, everyone has their taste and liking towards a type of music. Hard hitting evidence on why human brain reacts to music differently is not available but scientists have discovered some findings which state that the brain through cerebellum activation synchronizes the pulse of music with the neural oscillators. While processing music brain's language centre, emotional centre and memory centre are connected thereby stimulating a thrill obtained by expected beats in a pattern to provide a synesthetic experience [6].

This project was therefore aimed to provide people with befitting music using facial recognition, saving the time which is required to go into the files and scroll at a never ending list of songs to choose from thereby enhancing user experience.

2. RELATED WORK
There is ample research and work carried out in the field of emotion detection from faces .Some relevant work which was surveyed were mainly industry standard and leading research in the field. Viola and Jones [4], proposed a robust way to detect human faces in the image. They introduced a new image representation known as "Integral Image". They implemented a simple and efficient classifier by using Adaboost learning algorithm to select a small number of critical features from large set of features. Lastly, they introduced a method for combining classifiers in a cascade like structure which quickly discards background regions of the image.

Another popular technique for face detection is the classic Histograms of Oriented Gradients (HOG). Deniz, Bueno, Salido, and Torre [7] contributed towards, firstly to minimize

the errors occurring during the process of face detection due to partial occlusion, pose, and effect of brightness. Secondly it captures major structure for face detection. Thirdly they worked on reduction of noise removal from images for effective classification. Dalal and Triggs [8], in their paper reviewed existing and gradient based descriptors, they showed experimentally that grids of Histograms of Oriented Gradient descriptors significantly outperform existing feature sets for human detection they concluded that fine-scale gradients fine orientation binning, relatively coarse spatial binning, an high quality local contrast normalization in overlapping descriptor blocks are all important for good results.

For feature detection in detected face Visutsak[9], proposed an image based approach for emotion classification through lower facial expression. He used A-SVM classifier to classify the features in seven different emotions namely neutral, disgust, happy, sad, angry, fear, and surprised. Kazemi and Sulivan developed a efficient technique for landmark extraction from detected face in there paper One Millisecond Face Alignment with an Ensemble of Regression Trees [10]. They addressed issue of effectively estimating the face's landmark position. The landmark positions are the important points on the face which will be consider as feature for classification. They showed an ensemble of regression trees which can be used to estimate the face's landmark position directly from a sparse subset of pixel intensities. This technique gives very high quality predictions.

For classification of feature points Dumas [11], showed that SVM can be applied to the problem of classifying emotions on human faces. Her experiments showed that performance of SVM was equivalent to the performance of neural network.

3. OVERVIEW OF METHOD

The overall system can be divided into various logical stages like capturing image, detection of face, detection of landmark points on the detected face, classifying those feature points with the help of SVM classifier, and then generating the playlist according to that recognized mood. During the training phase a dataset of images will be created to train the SVM classifier while after implementation of the system a single captured image can be given to trained SVM file to predict the mood.

The different moods in which the system will classify the images are happy, neutral, and sad. The system will pre-sort the songs according to their genre in the above mentioned categories. Using these presorted lists and some heuristics like the most frequently played songs, personalized playlists will be created.

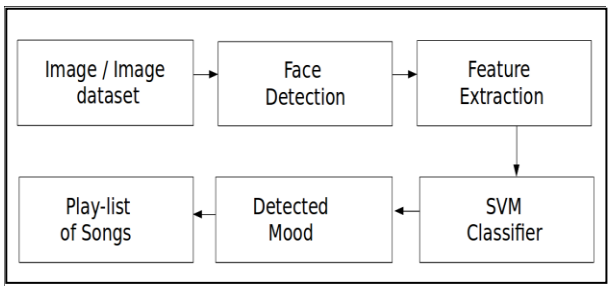

Fig 1: Block Diagram of MoodyPlayer

3.1 Database creation

In order to gain high accuracy in correctly predicting the mood, properly training model is a vital step. To have a good trained model giving a dataset which accurately justifies the generalized nature of human face is also important. The dataset must consist of images of subject or person with varied face structure and facial expressions. If the system recognizes and processes different ethnically differentiating facial structure it will more generic in nature and system can be effective generalized with high performance.

For proposed system to work, dataset must contain images of human frontal faces only. Dataset contains only faces of adult male and female subjects having age range between 18 to 36 years portraying different emotions. Based on the above criteria, selection was done from multiple databases which were Karolinska Directed Emotional Faces *(*KDEF) [12], Taiwanese Facial Expression Image Database [13], and The Japanese Female Facial Expression (JAFFE) [14] Database. Along with these an additional database is also created which is named as Indian Facial Emotion Expression Database (IFEED) and collectively all this forms a single dataset which has been categorized into three different emotion i.e. happy, neutral, sad, and they are accordingly labeled.

Conventionally dataset is divided into training dataset and testing dataset using some thumb rule but proposed system will use K-fold cross validation. To build a generic dataset certain criteria were imposed while selecting dataset these criteria were 1) Healthy ratio of male and female, in which there are 80 male and 102 female images. 2) Diversity in age i.e., a broad age range from 18 to 36 is chosen. 3) Care is also been taken to represent a good amount of subject from varied geographical region i.e., Indian, Japanese, South Asian, South American. After applying the above criteria's, the dataset contains 1050 images in total with 350 in each of the three moods.

3.2 Face detection

It is important for an image to be well lit and recognizable for the system to detect face. The system should read faces in low light but if the light is below certain accepted limit the system gives a prompt to retake as no face is detected.

Image is divided into the fixed size blocks. The blocks from which Histogram of Gradient (HOG) features are extracted and are overlapped on the detector window, the extracted vectors are fed as input to linear SVM in order to classify object or non-object classification and obtain result as the extracted face.

3.3 Feature extraction

When the system takes the images as an input it's very important to device an algorithm to detect emotions. This detection of emotions can be obtained by extracting the feature of the face for example the lips are stretched and make an extended 'v' type structure it means person is very happy similarly if eyebrows are arched and lips are in the shape of small 'n' then it amounts to a sad emotions. Number of permutations and combinations like the above stated determine the appropriate emotion through face and classify them into different present emotion categories.

Feature extraction is necessary stage for predicting the correct mood of the user so proposed system will use ensemble of regression trees to obtain facial landmark positions from sparse subset of pixel intensities. For robust performance with high quality prediction gradient boosting for training an

ensemble of regression trees is used to minimize the misclassified and partially labeled data.

There are total 52 landmark points on the face that system will detect and these points will be the data on basis of which the mood will be predicted. These extracted feature points are further passed to the SVM for classification.

3.4 Classification

Classification is a general process related to categorization. It is the action or process of clustering something. System is capable to classify images into different emotions. For classification of images, system will use Support Vector Machine (SVM). There are two phases in preparing an efficient classifier.

3.4.1 Training Phase:

The training dataset is divided into three classes for three moods i.e. happy, neutral, and sad , for each of these corresponding mood values +1 ,+2, and +3 is assigned in the training file as the correct label for this model to train. Here giving feature point along with its correct label is at most important as the learning of the model is supervised learning. The feature point are the position of the landmark on the face which were calculated in the previous step of feature extraction these feature point are converted in a single dimension from a two dimension X,Y coordinate system.

3.4.2 Testing Phase:

In testing phase , a similar input is given to the train model as it was in training phase with an exception that correct label is not known by the system for predicting the mood. It is just use to check the accuracy of the predicted result.

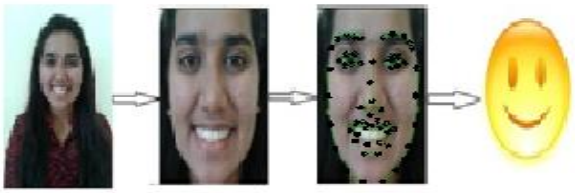

Fig. 3: Mood detection phases of MoodyPlayer

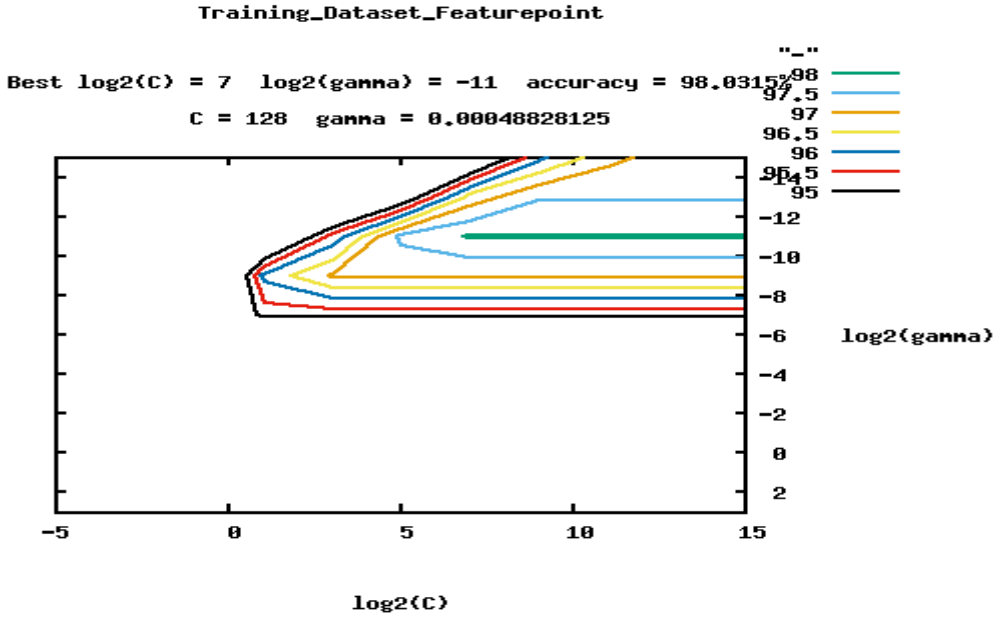

Fig. 4: Optimized parameter for C-SVM with RBF kernel

The figure depicts system's overview. It processes image from IFEE database, detecting the face using the HOG technique, then detecting the facial landmark points and classifying it using SVM classifier to predict the mood happy.

3.5 Music classification

In this stage, the predicted mood of the user is used to classify songs and to create a playlist for the particular mood. Songs are categorized into three groups representing the previously specified mood. This classification of songs happens on the basis of genre of the songs. For experimentation only English songs are considered which were freely available to download and use. These songs were manually categorized into above said three groups.

4. EXPERIMENT AND RESULTS

K-fold cross validation is the process in which the original dataset is divided randomly into proportionate number of k sub datasets. Out of k sub datasets, k-1 sub-datasets are used as training data for training the model and the remaining sub-dataset is used for validating the same model. The above process is repeated k times and one of the unique sub-dataset is used for validation and the remaining k-1 for training the model. The results of all the iteration is average and this result is the final accuracy of the system.

To have an appropriate number of images in training dataset a suitably chosen value of k as 5 is used. Before selecting C-SVM and Radial Basis Function (RBF) as a kernel in the system, multiple experiments with C-SVM and Nu-SVM with different combinations of kernels were carried out. Optimized value of parameters were fed to the model. The dataset for training and testing was comprised of standard databases like KDEF [12], TFEID [13] and JAFFE [14]. And a dataset of human faces which comprised specifically of Indian faces. Through this experimentation best results were yielded from the combination of C-SVM with Radial Basis Function Kernel. For testing the performance of the system on Indian faces separate dataset was created specifically of Indian faces only.

Table 1: Results of classification for various SVM kernel type

SVM-Type	Kernel Type	Accuracy (%)	
		Indian Dataset	Standard Dataset
C-SVM	1.Linear	75.33	80
	2.Polynomial	73.83	87.5
	3.RBF	**81.5**	**89.5**
	4.Sigmoid	33.33	40
Nu-SVM	1.Linear	73	80
	2.Polynomial	66	80
	3.RBF	75	80
	4.Sigmoid	33.33	20

To carry out this experiments, Dlib[15] and LIBSVM[16] libraries were used.

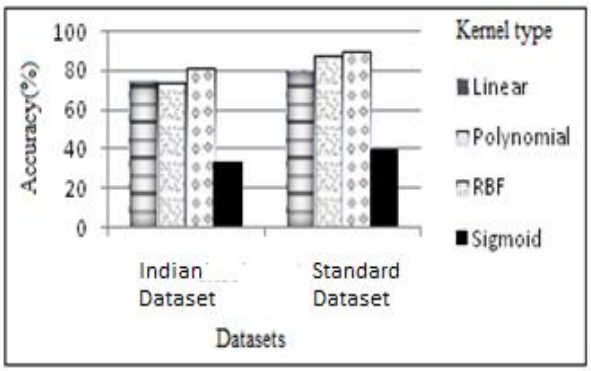

Fig 5: Accuracy plot of C-SVM for Indian and Standard datasets

Dlib library was used for face and facial landmark detection. These detected points were classified using SVM classifier which was created with the help of LIBSVM library.

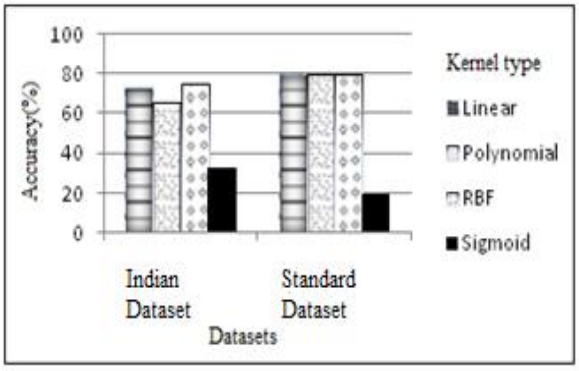

Fig 6: Accuracy plot of Nu-SVM for Indian and Standard datasets

5. CONCLUSION AND FUTURE SCOPE

We have designed MoodyPlayer a mood based music player which uses a face detector which is based on Histograms of Oriented Gradients (HOG) descriptors technique and a facial landmark detection system. Further, this facial landmark points were classified into three different moods using C-SVM with Radial Basis Function kernel to have an overall accuracy of 89.5%, which was calculated using K-fold(K=5) cross validation method. The system is able to effectively categorize the songs based on the detected mood.

In future MoodyPlayer can be enhanced with the capability of detecting the mood of a group rather than individuals. And can be than effectively use in public places and gatherings. The system with some additional functionality can act as a mood lifter or mood enhancer.

6. ACKNOWLEDGMENTS

The authors would like to thank esteemed Head of Department Prof. S. P. Patil for his valuable support and encouragement.

7. REFERENCES

[1] G. Heamalathal, C.P. Sumathi, A Study of Techniques for Facial Detection and Expression Classification, International Journal of Computer Science & Engineering Survey(IJCSES) Vol.5, No. 2, April 2014.

[2] Bernhard Fröba Christian Küblbeck Real-Time Face Detection Using Edge-Orientation Matching AVBPA '01 Proceedings of the Third International Conference on Audio- and Video-Based Biometric Person Authentication Pages 78-83 Springer-Verlag London, UK ©2001 table of contents ISBN:3-540-42216-1

[3] Oliver Jesorsky, Klaus J. Kirchberg, and Robert W. Frischholz BioID AG, Berlin, Germany, Robust Face Detection Using the Hausdorff Distance. Third International Conference on Audio- and Video-based Biometric Person Authentication, Springer, Lecture Notes in Computer Science, LNCS-2091, pp. 90–95, Halmstad, Sweden, 6–8 June 2001.

[4] PAUL VIOLA Microsoft Research, One Microsoft Way, Redmond, WA 98052, USA And MICHAEL J. JONES Mitsubishi Electric Research Laboratory, 201 Broadway, Cambridge, MA 02139, USA Robust Real-Time Face Detection International Journal of Computer Vision 57(2), 137–154, 2004_c 2004 Kluwer Academic Publishers. Manufactured in The Netherlands.

[5] http://changingminds.org/techniques/body/parts_body_la

nguage/face_body_language.ht

[6] The psychological functions of music listening Christine Städtler and David Huron US National Library of Medicine National Institutes of Health Published online 2013 Aug 13. Prepublished online 2013 May 24.

[7] Face recognition using Histograms of Oriented GradientsO. Déniz, G. Bueno, J. Salido, F. De la Torre Universidad de Castilla-La Mancha, E.T.S. Ingenieros Industriales, Avda. Camilo Jose Cela s/n, 13071 Ciudad Real, Spain Carnegie Mellon University, Robotics Institute, 211 Smith Hall, 5000 Forbes Ave., Pittsburgh, PA 15213, USA Pattern Recognition Letters 32 (2011) 1598–1603

[8] Histograms of Oriented Gradients for Human Detection Navnee Dalal And Bill Triggs INRIA Rhone-Alps,655 Avenue de l' Europe, Montbonnot 38334, France

[9] Porawat Visutsak Emotion Classification through Lower facial Expressions using Adaptive Support Vector Machines

[10] Vahid Kazemi and Josephine Sullivan One millisecond face alignment with an Ensemble of regression Trees Computer Vision Foundation CVPR2014

[11] Melanie Dumas Department of computer science,University of california,san Diego CA 92192-0114

[12] Lundqvist, D., Flykt, A., & Öhman, A. (1998). The Karolinska Directed Emotional Faces -KDEF, CD ROM from Department of Clinical Neuroscience, Psychology section, Karolinska Institutet, ISBN 91-630-7164-9.

[13] Li-Fen Chen and Yu-Shiuan Yen.(2007). Taiwanese Facial Expression Image Database. Brain Mapping Laboratory, Institute of brain Science, National Yang-Ming University, Taipei, Taiwan.

[14] Michael J. Lyons, Shigeru Akemastu, Miyuki Kamachi, Jiro Gyoba.Coding Facial Expressions with Gabor Wavelets, 3rd IEEE International Conference on Automatic Face and Gesture Recognition, pp. 200-205 (1998).

[15] Davis E. King. Dlib-ml: A Machine Learning Toolkit. Journal of Machine Learning Research 10, pp. 1755-1758, 2009

[16] C. Chang and C.-J. Lin. LIBSVM : a library for support vector machines. ACM Transactions on Intelligent Systems and Technology, 2:27:1--27:27, 2011.