A Music Visual Interface via Emotion Detection Supervisor

Sourabh Rakhunde, Ritika Mathur, Kashih Shah, Pushpendra Rajguru

Computer Engineering, Pune University Sinhgad Institute Of Technology, Lonavala

Abstract: *For a particular emotion of the user, the System evaluates songs according to the qualification assessed by two factors which are: song's relevancy to the user's preference, and song's mental influence on the user's feeling. In this proposed system, user's emotion is not input manually by the user, but detected automatically by the machine. In order to do that, user's facial expression data is captured from the webcam and then used as inputs for emotion detecting process. The motivation behind this system is the lack of a context-aware Music Recommendation System where automatically detected user's mood plays the most important role as a contextual key. The need of such system is made obvious by the fact that digital music libraries are constantly expanding, which thus makes it remarkably difficult for listeners to recall a particular song matching their present mood.By training the system to recognize user's emotional state by facial expression, it is made possible for listeners to generate a playlist which suits with their current emotion, and of which songs are rated also by the potentially mental influence on user's emotion.*

Keyword:*Face recognition, Face detection, PCA, Emotion Extraction and detection, Euclidean Distance.*

I.   Introduction

It is undeniable that music plays a very important role in human's life. The storage of those portable music devices is continuously increasing and now has reached to the number of 160GB, allowing users to load thousands of songs into their libraries. However, besides the convenience, that characteristic on the other hand troubles the users as they may experience the confusion while choosing songs. For that reason, Music Recommendation System becomes a very essential subject in current context. So, the main purpose of behind the development of this system is to provide users an efficient Music Recommendation System that automatically detects users' preference & generates a playlist which suits with their current emotion, and of which songs are rated also by the potentially mental influence on user's emotion. Moreover, the core logic will reside in sequence of reusable web services. Other users or application can use web service of face recognition and emotion recognition.

II.   Objective

1. Person Independent for identifying Emotion- Any person using this application will get the desired output that is his emotion (impromptu) will be recognised without his or her effort.
2. Develop a Compatible music player- A compatible, easy to use music player will be created playing music accordingly to the emotion/mood extracted of the user.
3. To create reusable application software- This application follows the principle of reusability. It can be reused multiple times and the result detected will be fruitful every time due to the PCA algorithm.
4. To recommend music based on user's present mood by detecting the emotion- Once the emotion is extracted out, the mood of the user is detected (happy, angry, sad, afraid, excited, surprised etc.).
5. To identify Emotion comprehensively- Principal Component Analysis is used to extract the features of the face that is through lips, nose, eyes, eyebrows, cheeks etc.

III.   Principal Component Analysis

Principal Component Analysis (PCA) is a dimensionality reduction technique based on extracting the desired number of principal components of the multi-dimensional data. The purpose of PCA is to reduce the large dimensionality of the data space (observed variables) to the smaller intrinsic dimensionality of feature space (independent variables), which are needed to describe the data economically. This is the case when there is a strong correlation between observed variables. The first principal component is the linear combination of the original dimensions that has the maximum variance; the nth principal component is the linear combination with the highest variance, subject to being orthogonal to the n -1 first principal components.The assumption relates to the Eigen spectrum - the set of the eigenvalues of the data covariance matrix. The I-th eigenvalue is equal to the variance along the i-th principal component; thus, a reasonable algorithm for detecting k is to search for the location along the decreasing Eigen spectrum where the value of l drops significantly. Since the basis vectors constructed by PCA had the same dimension as the input face images, they were named "Eigen faces". PCA is an information theory approach of coding and decoding face images may give insight into the information

content of face images, emphasizing the significant local and global "features". Such features may or may not be directly related to face features such as eyes, nose, lips, and hair.

These eigenvectors can be thought of as a set of features that together characterize the variation between face images. Each image location contributes more or less of each eigenvector; so that we can display the eigenvector as a sort of ghostly face which we call an *Eigen face*. Each individual face can be represented exactly in terms of a linear combination of the Eigen faces. Each face can also be approximated using only the "best" Eigen faces-those that have the largest eigenvalues and which therefore account for the most variance within the set of face images. The best M Eigen faces span an M-Dimensional subspace- "face space" − of all possible images. This approach of face recognition involves the following initialization operations:

1. Acquire an initial set of face images (the training set).
2. Calculate the Eigen faces from the training set, keeping only the M images that correspond to the highest eigenvalues. These M images define the face space. As new faces are experienced; the Eigen faces can be updated or recalculated.
3. Calculate the corresponding distribution in M-dimensional weight space for each known individual, by projecting his or her face images onto the "face space".

Having initialized the system, the following steps are then used to recognize new face images:
1. Calculate a set of weights based on the input image and the M Eigen faces by projecting the input image onto each of the eigenfaces.
2. Determine if the image is a face by checking to see if the image is sufficiently close to "face space".
3. If it is a face, classify the weight pattern as either a known person or as unknown.

By concatenating each row of the image by row, a face image can be transformed to a column vector. Assume the width and height of the image are n pixels and m pixels
respectively, the size of the transformed vector of this image will be n × m by 1. Given M
facial expression images as training data, we convert these images to corresponding column image vectors Γ_i,

where i = 1, 2, …, M. Compute the mean of training data $\Psi = \dfrac{1}{M}\sum_{n=1}^{M}\Gamma_n$ and let the normalized vectors be $\Phi_i = \Gamma_i - \Psi$. We want to seek a set of M orthonormal vectors, un, that best represents the distribution of the data. The kth vector,u_k, is chosen such that

$$\lambda_k = \frac{1}{M}\sum_{n=1}^{M}(u_k^T \Phi_n)^2$$

(1)

is maximum where, $u_l^T u_k = \begin{cases} 1 & \text{if } l = k \\ 0 & \text{otherwise} \end{cases}$. Extend Eq. (1) and multiply u^T to both sides of the

equation, we obtain $\lambda_k u_k^T = u_k^T\left(\dfrac{1}{M}\sum_{n=1}^{M}\Phi_n\Phi_n^T\right)u_k u_k^T$. Assume the covariance matrix C = AA^T, where A =

$$\lambda_k u_k^T = u_k^T C.$$

[Φ1, Φ2, …, ΦM], we obtain   (2)

Transpose both side of Eq. (2), we derive $\lambda_k(u_k^T)^T = C^T(u_k^T)^T = C^T u_k$ and $C^T = AA^T = C$. Then we can conclude that $\lambda_k u_k = Cu_k$. Thus the vectors u_k and scalars λ_k are the eigenvectors and eigenvalues, respectively, of the covariance matrix C = AA^T.
Because the size of covariance matrix C is nm × nm, it is time-consuming to determine nm eigenvalues and eigenvectors. It is necessary to reduce the complexity of the computation. Define a matrix C ′ = A^{T}A, then the size of C ′ is M by M and let vi denote the eigenvectors of the matrix C ′, we obtain A^{T}Avi = μivi. Multiply A to both sides, we have

$$AA^T A_{vi} = \mu i A_{vi}. \qquad (3)$$

We can observe from Eq. (3) that A_{vi} is the eigenvector of C = AA^T. So we can now reduce the method by derive p (p < nm) eigenvectors (vi) of the matrix C ′ and derive the p eigenvectors of covariance matrix C by multiplying A to vi. The face images can be represented in the way by projecting the data in the image space

onto the face space (also called Eigenfaces). The dimension of the projected data in the feature space is much smaller than that in the original image space. After the Eigenfaces are obtained from the training set, we may transform face image into feature space by a simple projection operation $x = u^T (\Gamma - \Psi)$, i = 1, 2, …, p, where xi denotes the projection of the face image Γ projected onto the i-th Eigenface component ui, where Ψ is the average face image ofthe training set. The projections constitute a vector XT = [x1, x2, …, xp] called projectionvector. We will treat the projection vector X as the features for further training process.By measuring the similarity (mean-square error) between input image and the reconstructed image of each class, we can identify the class of input image whose reconstructed image is most similar to the input one.

IV.   A generic face recognition system

The input of a face recognition system is always an image or video stream. The output is an identification or verification of the subject or subjects that appear in the image or video.
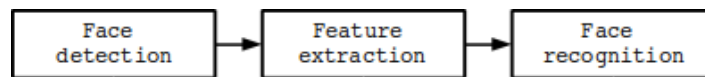

Figure 1: A generic face recognition system.

Face detection is defined as the process of extracting faces from scenes. So, the system positively identifies a certain image region as a face. These features could be certain face regions, variations, angles or measures, which can be human relevant (e.g. eyes spacing) or not. Finally, the system does recognize the face. In an identification task, the system would report an identity from a database. Face detection must deal with several well-known challenges. They are usually present in images captured in uncontrolled environments, such as surveillance video systems. These challenges can be attributed to some factors:

- **Pose variation** - The ideal scenario for face detection would be one in which only frontal images were involved. But, as stated, this is very unlikely in general uncontrolled conditions. Moreover, the performance of face detection algorithms drops severely when there are large pose variations. It's a major research issue. Pose variation can happen due to subject's movements or camera's angle.
- **Feature occlusion -** The presence of elements like beards, glasses or hats introduces high variability. Faces can also be partially covered by objects or other faces.
- **Facial expression -** Facial features also vary greatly because of different facial gestures.
- **Imaging conditions -** Different cameras and ambient conditions can affect the quality of an image, affecting the appearance of a face.

V.   Face detection problem structure

Face Detection is a concept that includes many sub-problems. Some systems detect and locate faces at the same time, others first perform a detection routine and then, if positive, they try to locate the face. Then, some tracking algorithms may be needed.
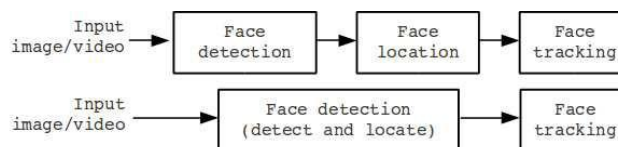

Figure 2: Face detection processes.

Face detection is, therefore, a two class problem where we have to decide if there is a face or not in a picture. This approach can be seen as a simplified face recognition problem. Face recognition has to classify a given face, and there are as many classes as candidates.

VI.   4.5 Feature Extraction

What it's clear is that we can recognize people we know, even when they are wearing glasses or hats. We can also recognize men who have grown a beard. It's not very difficult for us to see our grandma's wedding photo and recognize her, although she was 23 years old. All these processes seem trivial, but they represent a challenge to the computers.
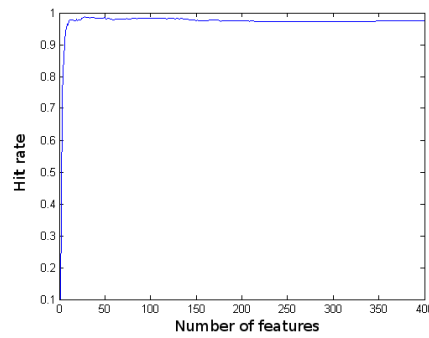
Figure 3: PCA algorithm performance

The feature extraction process must be efficient in terms of computing time and memory usage. The output should also be optimized for the classification step.

Feature extraction involves several steps - dimensionality reduction, feature extraction and feature selection. This step may overlap, and dimensionality reduction could be seen as a consequence of the feature extraction and selection algorithms. Both algorithms could also be defined as cases of dimensionality reduction. Dimensionality reduction is an essential task in any pattern recognition system. The performance of a classifier depends on the amount of sample images, number of features and classifier complexity. However, added features may degrade the performance of a classification algorithm. This may happen when the number of training samples is small relative to the number the features.

This problem is called "curse of dimensionality" or "peaking phenomenon". A generally accepted method of avoiding this phenomenon is to use at least ten times as many training samples per class as the number of features. This requirement should be satisfied when building a classifier. The more complex the classifier, the larger should be the mentioned ratio. This "curse" is one of the reasons why it's important to keep the number of features as small as possible. The classifier will be faster and will use less memory.
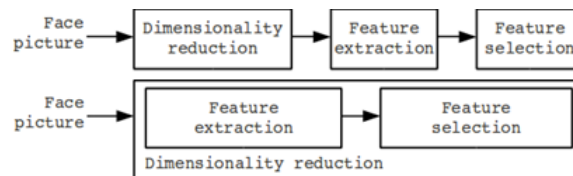

Figure 4: Feature extraction processes.

VII. Emotional Model

Emotional model is a graph representing specific emotions and their relationships in a regarded concept. Based on the approaches, emotional model can mainly be divided into two major types which are categorical and dimensional ones. A categorical emotional model is the model composed of several distinct emotional words representing classes which form the basis for other words 'classification. An example of remarkable study on categorical emotional model field is the work of researcher Paul Ekman. In order to propose a category of basic emotions, Ekman did a research on the Fore tribesmen of Papua New Guinea, who come from an isolated culture, testing their ability to identify emotions in photographs of people coming from cultures which are not familiar with them.And, he released a list of basic emotions. These emotions are manifested as human being's facial expression.
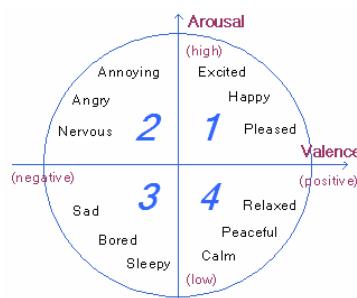

Figure 5: Emotion Graph.

VIII. Constraints

As per the research and techniques available for emotion detection, the perfection of the detected emotion is up to 65%.

Emotion recognition from music signal is a challenging task due to the following reasons:

- First, emotion perception is intrinsically subjective and people can perceive different emotions for the same song. This subjectivity issue makes the performance evaluation of an MER system fundamentally difficult because a common agreement on the classification result is hard to obtain.
- Second, it is not easy to describe emotion in a universal way because the adjectives used to describe emotions may be ambiguous, and the use of adjectives for the same emotion can vary from person to person.
- Third, it is still inexplicable how music evokes emotion. What intrinsic element of music, if any, creates a specific emotional response in the listener is still far from well-understood.
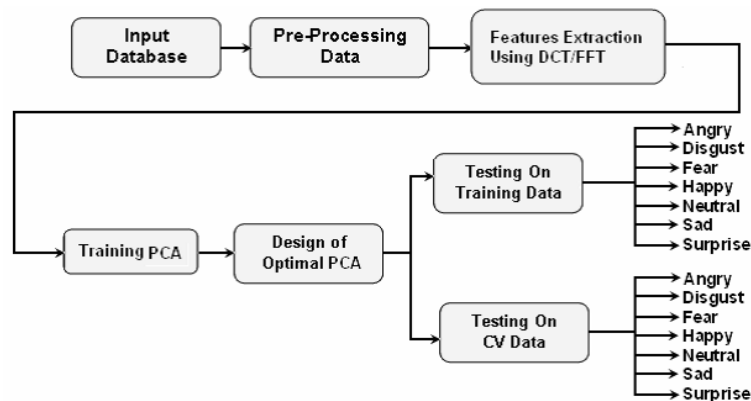

Figure 6: PCA working model.

IX. Experimental Results

Our goal is to come to a classification of 4 main emotional expression (i.e. Happy, Sad, Normal, Surprise) using human facial expressions. There are several algorithms available for that. We are trying to optimize those algorithms by applying image utility algorithm. We have generated a new optimized algorithm for emotion detection.

Table 1:

Actual → Predicated	Happy (%)	Sad (%)	Normal (%)	Surprise (%)
Happy	**91**	1	5	3
Sad	7	**78**	13	2
Normal	33	9	**53**	5
Surprise	40	5	13	**42**

X. Conclusion

In this paper, we have discussed how we can identify emotions through face detection and feature extraction using the PCA algorithm.

Thus, an interface between emotions and music can be created by playing the relevant playlist according to the detected emotion.

Reference

[1] A Regression Approach to Music Emotion RecognitionYi-Hsuan Yang, Yu-Ching Lin, Ya-Fan Su, and Homer H. Chen, Fellow, IEEE
[2] JOURNAL OF INFORMATION SCIENCE AND ENGINEERING,Facial Expression Classification Using PCA and Hierarchical Radial Basis Function Network, DAW-TUNG LIN, Department of Computer Science and Information Engineering
[3] Principal Component Analysis, A Powerful Scoring Technique, George C. J. Fernandez, University of Nevada - Reno, Reno NV 89557.
[4] Face Recognition: A Literature Survey, W. ZHAO Sarnoff CorporationR. CHELLAPPA, University of Maryland, P. J. PHILLIPS, National Institute of Standards and Technology and A. ROSENFELD,University of Maryland
[5] Face Recognition Algorithms, Proyecto Fin de Carrera, June 16, 2010, Ion Marqu´es
[6] Emotion detection in music, a survey Bram van de Laar

[7]   Performance Comparison of Principal Component Analysis-Based Face Recognition in Color Space Seunghwan Yoo1, Dong-Gyu Sim2, Young-Gon Kim1 and Rae-Hong Park11Sogang University,2Kwangwoon University,South Korea

XI.   Authors

1. **SourabhRakhunde,** Under Graduate, Sinhgad Institute of Technology, Lonavala,
2. **Kashish Shah**, Under Graduate, Sinhgad Institute of Technology, Lonavala,
3. **RitikaMathur,** Under Graduate, Sinhgad Institute of Technology, Lonavala,
4. **PushpendarRajguru,** Under Graduate, Sinhgad Institute of Technology, Lonavala,