Music Recommender System for users based on Emotion Detection through Facial Features

Ahlam Alrihaili
Department of Computer Science
Taibah University
iahlam9@taibahu.edu.sa

Alaa Alsaedi
Department of Computer Science
Taibah University
asaedi@taibahu.edu.sa

Kholood Albalawi
Department of Computer Science
Taibah University
Kalbalawi@taibahu.edu.sa

Liyakathunisa Syed
Department of Computer Science
Taibah University
drliyakathunisa@gmail.com

***Abstract*— In recent years, facial emotion detection received massive attention because of its applications in computer vision and human-computer interaction fields. Due to the active works in this field, various algorithms and applications were proposed and implemented. In this research, we propose a recommender system for emotion recognition that is capable of detecting the user emotions and suggest a list of appropriate songs that can improve his mood. A brief search was conducted on how music can affect the user mood in short-term to gain knowledge and enable us to provide the users with a list of music tracks that work well on improving the user moods. The proposed system detects the emotions, if the subject has a negative emotion then specific playlist will be presented that contains the most suitable types of music that will improve his mood. On the other hand, if the detected emotion is positive, a suitable playlist will be provided which includes different types of music that will enhance the positive emotions. Implementation of the proposed recommender system is performed using Viola-Jonze algorithm and Principal Component Analysis (PCA) techniques, we were able to implement the proposed system successfully in MATLAB(R2018a).**

Keywords—*Emotion Detection, Principal Component Analysis, Recommender System , Viola-Jonze.*

I. INTRODUCTION

Communicating between individuals is a major aspect of everyday life. It conveys accurate details and millions of information among humans, whether in the form of words, tone or expression. The expression provided by the face and body is the best means to understand people in communication. In particular, the facial expression is one of the basic elements of human communication and can be considered a form of nonverbal communication [1]. Facial expressions refer to a movement of one or more facial muscles or skin. The aim of these movements is to express the facial emotions of an individual such as happy, sad, natural and surprised.

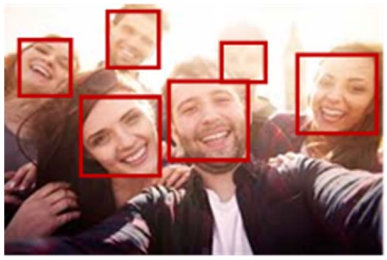

Figure 1. Face detection

Face detection and emotion detection are two separate concepts that are considered are given high significance in recent research fields like digital image processing, pattern recognition communities, and computer vision. Face detection is defined as the process that is composed of two steps: first, finding faces in a picture or video and does not care if the face refers to a person or not. Then, it draws a surround box of every face in the image in any expression, orientation, facial pose, illumination, and occlusions, shown in figure 1. Emotion detection is considered as one of the facial expressions and is the main process for identifying emotions of human's face.

Nowadays, emotion detection is considered as one of the most important techniques that are used in many applications such as smart card application, surveillance, image database investigation, criminal, video indexing, civilian applications, security and adaptive human-computer interface with multimedia environments [2]. Moreover, emotion detection provides more reliable and inexpensive way to know the opinions of customers about their products. In the past, most enterprise and shops rely on using traditional marketing methods such as advertising, sale points, satisfaction surveys, and price etc. These methods may be useful in some cases, but it is expensive, consumes a lot of time and are potentially unreliable.

Hence, the main objective of this paper is to implement a system that is able to detect user emotions, which can be happy, sad, natural or surprised. Then after the emotion is determined the proposed system will provide the user with a music playlist that contains music clips of certain music types which improve the user's mood. For example, if the user expression is classified as happy then the most suitable playlist will be enabled which is the classical music playlist to choose from various music clips. For the system implementation, we mainly depend on Viola-Jonze Algorithm and PCA to detect the user emotions.

The rest of the paper is organized as follows: section II presents the preliminaries. Literature review is presented in section III. Section IV presents the description of the proposed system. Results and discussion are presented in section V. Finally we conclude the paper in section VI.

978-1-7281-3021-7/19/$31.00 ©2019 IEEE
DOI 10.1109/DeSE.2019.00188

II. PRELIMINARIES

A. Viola-Jones algorithm

This algorithm was proposed by Paul Viola and Michael Jones in 2001 [3]. It is a Robust, fast and real-time object detection and used to detect different kinds of objects but it was highly motivated in use on detection of facial features like the mouth, eyes, face, and nose. The algorithm mainly depends on using the haar features selection filters. The haar feature selection filters means all human faces share a group of similar properties, for example, the human nose is brighter than the eye and eyes appear darker than the upper cheeks. Also, the place and dimension of the nose, eyes, and mouth similar in all human beings for example in figure 2, Haar feature show the same bridge of nose. Haar features differ in height and width based on the sum of the black and white pixel as shown in figure 3. The Haar features are able to detect different parts of the face such as the cheeks, nose, and eyes etc. The adaboost feature is used to remove unwanted features. Hence, based on Haar features and Adaboost feature face can be detected and resized on the standard resolution.

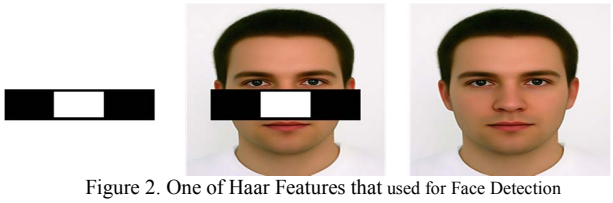

Figure 2. One of Haar Features that used for Face Detection

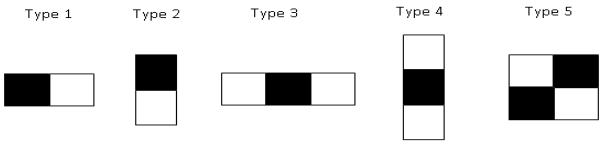

Figure 3. haar feature

B. Principal Component Analysis (PCA) method

This algorithm was proposed by Karl Pearsonin in 1901[4]. PCA is a statistical method which is based on orthogonal transformation that transform a set of observations of possibly correlated variables to a set of linearly uncorrelated variables of values called principal components [18]. PCA is used in data set to reduce the dimensionality and at the same time maintain the characteristics of dataset. Principal component analysis algorithms are used in a lot of applications and one of these applications is facial recognition and emotion detection.

The aim of PCA in emotion detection field is to convert each image in a training set to eigenface as shown in the following steps:

1. Collect set of images (training set) and all these pictures should have the same lighting conditions and aligned mouth and eyes among all images(normalized).
2. Convert each image in training set to one vector by concatenating the pixels of the rows and store it in a single column.
3. Store all images in the training set into one matrix. Each column represents an image.
4. Calculate the mean and then subtract it from each image in the training set.
5. Calculate the eigenvectors and eigenvalues of the matrix of covariance.
6. Select the principal components and arrange eigenvectors and sort the eigenvalues in descending order.

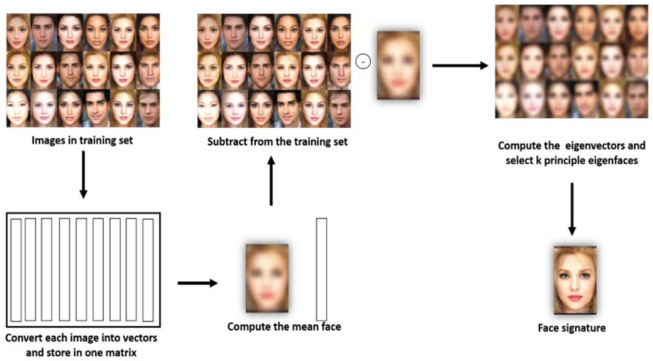

Figure 4. step of PCA method

III. LITERATURE REVIEW

A. Facial emotions Detection Techniques

Facial emotions is considered as the most important factor in people's communication which enable us to perceive other people's intentions. Typically, people deduce the other people emotional states such as joy, anger, sadness, and fear through two main channels: voice tone and facial expression [1]. Mehrabian [5], state that two-thirds of people communication is carried in non-verbal communication and facial expression represents the largest component in this percentage. While only one third of emotion is conveyed in the verbal communication. For that reason, nowadays one of the main fields in computer vision is facial emotion recognition due to its academic significant and commercial potential.

If we look at the studies in this field since its inception so far, we find that the methods used by researchers can be divided into two main classes: conventional and neural networks-based methods.

In conventional methods, to infer the human emotions from the still images, researchers work on detecting the face and landmark (e.g. eyes, nose, and mouth) as a first step, then, they extract spatial features and end this process by classifying the expressions based on the extracted features to conclude and output the result.

One of the studies that is applied in the conventional approaches is Ghimire and Lee research [6], they depend on two kinds of geometric features that rely on 52 landmark

position and angle. Also, they used a pair of classifiers: Support Vector Machine(SVM) for feature vectors and multi class Adaboost during dynamic time warping.

Dhavalikar and Kulkarni [7] proposed a face recognition system that was implemented in three steps: the first step involves detecting the face using different techniques: 1) YCbCr to detect skin color, 2) Lighting compensation, and 3) Morphological operations to select only the desired part of the face. In the second step, they extract the appearance features using AAM approach. Finally, the extracted features are recognized depending on Euclidean Distance approach. This system achieved a high accuracy rate estimated at 90%-95%.

The primary steps in any fundamental face recognition system consist of three steps: first, the face is detected, then face features extracted and represented as vectors, finally, the face is recognized. Based on that Agrawal and Khatri proposed a technique that employs the Viola-Jones system to detect faces in an image and the principal component analysis to extract features [8].

In another study [9], they introduced a system for face recognition that depends on the whole face image, not on the local features. They also utilized the Viola-Jonze for the face detection, PCA to extract the features from the images then the Linear Discriminant Analysis was applied over the eigen faces that are represented by the PCA to minimize the PCA limitation. Then the similarity between two images vectors were measured using the Square Euclidean Distance. They conclude in their study that as the images of the training increased , the rate of the recognition also improved.

The techniques that are used in the former two studies are similar to the techniques that we used in the implementation of the proposed system. We employed the Viola-Jones as a method for detecting the faces that are characterized by its capability in achieving advanced rate in recognizing faces. For features extraction and dimensionality reduction, PCA will be used.

In the meantime, the researchers went to use deep learning techniques that reduce in high percentage the dependent on face physics models and replace it by end-to-end learning models. There are many models for deep learning, but the Convolutional Neural Network (CNN) is the most widely used model in face recognition field.

Alizadeh and Fazel proposed a CNN for facial expression recognition to classify the expression into one of the seven well-known emotions types: happy, sad, angry, neutral, fear, surprise, and disgust. They conclude that CNN can successfully learn the facial features and detect facial emotions in an improved manner. In addition, they found that using the hybrid features did not add any improvement for the model accuracy, and they recommend to depend only on one raw pixel data for model learning purpose [10].

B. Studies on Facial expression recognition Applications

Recently, there has been considerable interest from the commercial sector, especially the retail sector, in using the technology of facial expression recognition to measure the satisfaction of its customers.

The well-known Walmart retails-Stores in America lately announced its intention to build a facial recognition system to measure in real-time their customer sentiment and take the appropriate action to improve the customer experience and to preserve their loyalty in the long term [11].

CaliBurger restaurants in California is another example on the retail sector that employed facial recognition technology by linking it to the loyalty program. They installed the software in their ordering kiosks, and the program works on to identify the customer's faces who registered in the company's loyalty program and then display a menu relying on customer preferences. John Miller, CEO of Cali Group state that the software application will help the restaurant chain to provide an interactive, specialized experience [12].

Karim et. al., proposed system that was able to successfully produce statistical information about customer satisfaction, they used image processing approaches to implement the system. Their system is able to identify regular customers faces, blacklisted customers, expression classification, and for new customers, they used age and gender classification. They conduct the experiments in a coffee shop at AIT (Asian Institute of Technology). In the expression module, they used Levi and Hassncer [13] CNN model to identify the emotions into three types: positive, negative, and natural. While the happiness will be recognized as a positive emotion, and anger, disgust, sadness, and contempt classified as negative emotions [14].

C. Music and their effect on the Mood

Different studies investigated the music and how it will affect the human in different ways, some researchers studied how music affects people modes in the short term or long term, while other focused on its impact on customers' behavior and measured their satisfaction on the service provided [15], Also some researchers proved its usefulness in patients' therapy [16].

Since the main concern in the proposed system centered around detecting the user emotions and improving their moods by music, we are going to deal with four types of emotions or moods, and we have to choose four types of music that best suit it.

In literature, it has been proven that the Classical music can reduce the anxiety and depression in short-term effectively, while, new age music that characterized by loud music, and fast tempo, this type of music increase relaxation and decrease tension. On the opposite side, listing to the Grunge Rock increases the negative emotions and decrease positive ones. Designer music that was identified by the upbeat rhythm will work on increasing emotional and mental balance. For the happy emotion, we intend to emphasize this emotion, so we compose the happy playlist of classical music

tracks, while the natural playlist consists of new age music and the surprised and sad playlist of designer music tracks [17].

IV. Proposed Methodology

The proposed system is a music controller that is based on automatic emotion detection. A webcam is used to capture the images that will be used as input to the proposed system, then it goes to the expression detector to classify it to one of four classes "Happy", "Natural", "Sad", and "Surprised" as shown in Figure 5.

Figure 5. Examples on the Four Emotion

Depending on the expression or the emotion that are detected the corresponding music playlist will be presented to the user to select from a group of the music clips. For example, if the user expression is classified as happy then the happy music playlist will be enabled to the users to choose from the music clips. For the dataset, we used free dataset that is available on the internet, it contains 182 images in four emotions: happy, natural, surprised, and sad. The stages of the proposed approach is shown in Figure 6.

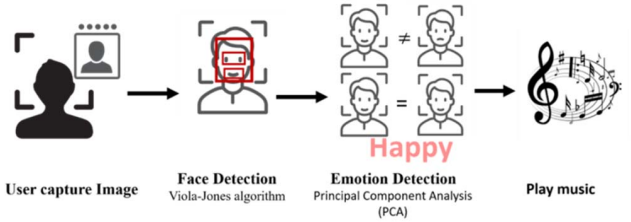

Figure 6. The proposed System Architecture

A. Image Acquisition

As a first step in the proposed system, we begin by acquiring the image of the user's face using a built-in laptop webcam (or any external camera can be employed). The face image to be correctly processed in the proposed system must contain one face in the frontal position in a uniformly illuminated background. Also, it should not be on the user face anything that could impede the process of detection such as glasses. For that, we allow the user to recapture the image to get a picture compatible with the standards of the proposed system in case the first picture does not fulfill the criteria. To enable that in MATLAB the support package of image acquisition toolbox for OS generic video interface was installed.

B. Face Detection

After acquiring the image, the system will start to detect the face by applying the Viola-Jones algorithm. This Algorithm is considered as one of the first frameworks that recognize the objects in real time. Simply, Viola-Jones scan the images using a sub-window to detect the features of the face in the image. When the face is determined, the image is cropped to contain the face only, to enhance the proposed system performance. Also, the Viola-Jones is reused to identify and crop the left and right eyes and mouth separately. The outcome of this step is four images, face, right eye, left eye, and mouth images.

C. Emotion Detection

Next the user sentiment must be detected, we use PCA method which is a well-known approach that is often used to detect face emotions. PCA will construct the face space and the eigenvectors that has the highest eigenvalues will be selected. Also, the acquired image will be projected over the face space. After that the emotion is detected by computing for the user image the scores for each emotion, then the emotion of the image is determined by getting the maximum score of the calculated emotion scores.

D. Enabling the correspondent Emotion playlist

The proposed system is depending on the detected emotion will present the correspondent music playlist. Since we have four emotions, we also have four playlists that offer music clips that are carefully chosen. For happy emotion, the classical music playlist will be activated, while the new age music playlist is dedicated to the natural emotion. For the negative emotions, surprised and sad the designer music playlist will be enabled to enhance the user mood to a better mood.

V. Result and Discussion

In the proposed work, we were able to successfully implement a system that utilizes the Viola-Jonze algorithm with the help of PCA to detect the emotion that are depicted in the acquired image to suggest depending on that the most suitable playlist will be recommended through which we aim to improve the user's mode. Figure 7 and 8 shows a screenshot of the proposed system and how the system provides the appropriate list based on the detected mode.

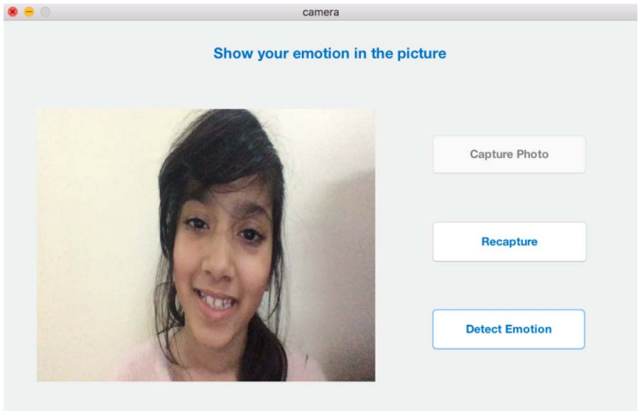

Figure 7. The system interface to capture the image.

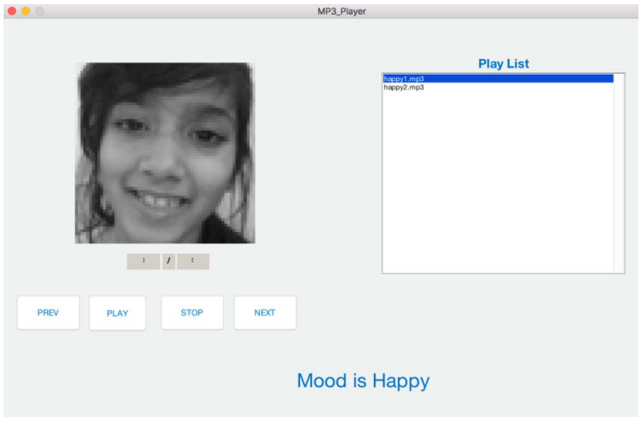

Figure 8. The result of emotion detection with matching music.

During the testing phase, we asked five people to test the proposed system and each one capture a single image for the four emotions: happy, natural, surprised, sad. As shown in table1 and figure 9, we found that most of the time the proposed system detects the emotions correctly. There are further observations we would like to highlight, we noticed that the happy face without showing the teeth may be classified as neutral, likewise, surprised face when the teeth appears may classified as happy, and when the teeth don't appear sometimes it's classified as sad because of the shape of the mouth and that may justify the cases that their detection was not accurate. The testing images are real-time images which affected the accuracy. The accuracy results in table1 depends on the training dataset, so if we increase the size of the training set the accuracy will be improved.

TABLE 1 ACCURACY OF EMOTION DETECTION IN THE PROPOSED SYSTEM

Person	Mode	Accuracy for the correct mode	Right Mode
Person 1	Happy	82%	Yes
	Sad	76%	Yes
	Neutral	97%	Yes
	Surprised	69%	Yes
Person 2	Happy	79%	Yes
	Sad	73%	Yes
	Neutral	98%	Yes
	Surprised	99%	Yes
Person 3	Happy	46%	No (Neutral with accuracy 72%)
	Sad	59%	Yes
	Neutral	98%	Yes
	Surprised	52%	No (Happy with accuracy 62%)
Person 4	Happy	60%	Yes
	Sad	45%	Yes
	Neutral	84%	Yes
	Surprised	60%	No (Sad with accuracy 0.072)
Person 5	Happy	40%	Yes
	Sad	66%	Yes
	Neutral	69%	Yes
	Surprised	56%	Yes

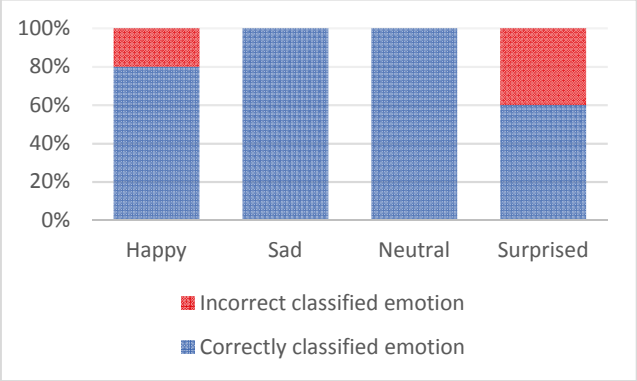

Figure 9. The image classification results.

VI. CONCLUSION

In this paper, we provide an overview about how the music can affect the user's mood and how choosing the right music tracks improve the user moods. Also, we showed the emotion detection techniques. To implement the proposed system, we used PCA and viola-Jonze Algorithm. The implemented system were able to detect the user emotions. The emotions that the system can detect were happy, sad, natural or surprised. After determining the user's emotion, the proposed system provided the user with playlist that contains music matches that detected the mood. The music was chosen to improve the user emotion after studying which type of music is suitable for each mood.

REFERENCES

[1] B. C. Ko, "A Brief Review of Facial Emotion Recognition Based on Visual Information," *sensors,* vol. 18, p. 401, 2018.

[2] "Principal component analysis." [Online]. Available: https://en.wikipedia.org/wiki/Principal_component_analysis. [Accessed: 29-Nov-2018].

[3] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, 2001, pp. I-I.

[4] M. A. Turk and A. P. Pentland, "Face recognition using eigenfaces," in *Computer Vision and Pattern Recognition, 1991. Proceedings CVPR'91., IEEE Computer Society Conference on*, 1991, pp. 586-591.

[5] A. Mehrabian, "Communication without words," *Communication theory,* pp. 193-200, 2008.

[6] D. Ghimire and J. Lee, "Geometric feature-based facial expression recognition in image sequences using multi-class adaboost and support vector machines," *Sensors,* vol. 13, pp. 7714-7734, 2013.

[7] A. S. Dhavalikar and R. Kulkarni, "Face detection and facial expression recognition system," in *Electronics and Communication Systems (ICECS), 2014 International Conference on*, 2014, pp. 1-7.

[8] S. Agrawal and P. Khatri, "Facial expression detection techniques: based on Viola and Jones algorithm and principal component analysis," in *Advanced Computing & Communication Technologies (ACCT), 2015 Fifth International Conference on*, 2015, pp. 108-112.

[9] N. H. Barnouti, S. S. M. Al-Dabbagh, W. E. Matti, and M. A. S. Naser, "Face Detection and Recognition Using Viola-Jones with PCA-LDA and Square Euclidean Distance," *International Journal of Advanced Computer Science and Applications (IJACSA),* vol. 7, pp. 371-377, 2016.

[10] S. Alizadeh and A. Fazel, "Convolutional Neural Networks for Facial Expression Recognition," *arXiv preprint arXiv:1704.06756,* 2017.

[11] A. Park, "Customer Intelligence: Facial Recognition Technology", *Inmoment.com*, 2018. [Online]. Available: https://www.inmoment.com/blog/how-facial-recognition-tech-will-lead-to-more-in-store-intelligence/. [Accessed: 29- Sep- 2018].

[12] R. Ferguson, "Facial recognition software comes to loyalty", *RetailWire*, 2018. [Online]. Available: https://www.retailwire.com/discussion/facial-recognition-software-comes-to-loyalty/?utm_source=Listrak&utm_medium=Email&utm_term=http%3a%2f%2fwww.retailwire.com%2fdiscussion%2ffacial-recognition-software-comes-to-loyalty%2f&utm_campaign=Today+on+RetailWire%3a+Alexa-enabled+glasses%3b+Kohl%27s+omni-progress%3b+Facial+recog+loyalty+-+1%2f9%2f18. [Accessed: 29- Sep- 2018].

[13] G. Levi and T. Hassner, "Age and gender classification using convolutional neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2015, pp. 34-42.

[14] N. T. Karim, S. Jain, J. Moonrinta, M. N. Dailey, and M. Ekpanyapong, "Customer and target individual face analysis for retail analytics," in *Advanced Image Technology (IWAIT), 2018 International Workshop on*, 2018, pp. 1-4.

[15] M. Mandila and V. Gerogiannis, "The effects of music on customer behavior and satisfaction in the region of Larissa–the cases of two coffe bars," in *International conference on contemporary marketing issues (ICCMI)*, 2012.

[16] C. L. Pelletier, "The effect of music on decreasing arousal due to stress: A meta-analysis," *Journal of music therapy,* vol. 41, pp. 192-214, 2004.

[17] R. McCraty, B. Barrios-Choplin, M. Atkinson, and D. Tomasino, "The effects of different types of music on mood, tension, and mental clarity," *Alternative therapies in health and medicine,* vol. 4, pp. 75-84, 1998.