Turkish Online Journal of Qualitative Inquiry (TOJQI)
Volume 12, Issue 7, July 2021: 8276 - 8281

Research Article

Expression Tuna – World of Expressive music

Anisetti Praharsha [a], Dr. Satya Prasad Lanka[b], Udayini Chandana [c]

[a]U.G. Scholar, [b]Professor, [c]Assistant Professor
[a,b,c] Stanley College of Engineering and Technology for Women, Hyderabad, Telangana
[a]anisettypraharsha@gmail.com , [b]principal@stanley.edu.in , [c] cudayini@stanley.edu.in

Abstract

Music forms an integral part of human beings and acts as a channel to help an individual to express his/her emotions. It also connects the listener with stimulus related to other emotions. The mood or emotion of the individual is communicated partly through the expressions on the face. Demand therefore exists for a system to assess the emotion or mood of a person and play the music to elicit the emotion. A technical approach includes the system recognizing the captured face expressions in a numeric form and identifying the individual's emotion determined by the centroid of these features. A musical piece or song is played according to this emotion from the play-list.

Keywords: Facial expression, Pattern vector, Emotion detection, music map, Image processing

1. Introduction

Most of the music lovers listen to the music reflecting their mood or feeling at a certain moment of time. Numerous traditional music players exist in which the listeners need to undergo the action of typing out their emotion or mood type, searching through the song-list and finally listening to it. During this process, they may lose interest as it is time-consuming and distracting. They may also undergo mood fluctuations. There is the facility of advanced features like recommending similar songs depending on the singer or genre offered by several music systems, but there is a lot of scope for improvement in this process through the automation of its different functions.

2. Basis of the work

The challenges related to face detection include that of non-frontal or partially visible or tiny faces. Also, the segmentation of the face from the background is a difficult task requiring the understanding of illumination and occlusion variations which can be resolved by considering texture [5]. The combination of the Histogram of Oriented Gradients (HOG) and Support Vector Machine (SVM) to calculate the detection ratio while using a high-quality image is discussed in detail but detection in actual practice is not possible [8]. According to this algorithm, an approach consisting of a simple sliding window is used for determining and representing the image in a pyramidal structure. Multiple scaling factors are used in this process. A step size of constant value is used to obtain small sub-images of 64 x 128 sizes and at different scales. Each of the sub-images is used to check if it is part of the face or not, and its HOG is computed. This data is next passed to the classifier (Linear or not) to make a decision of whether the sub-image represents a Face or not. After obtaining the resulting information in the form of pyramid, the operation of non-maxima suppression (NMS) is executed to discard the non-relevant information in the form of stacked rectangles. The superposition of one expression over the other is considered, which represents the blended expression [3]. A number of complex methods, like Neural Networks, Direction Intensity Signature and kernel based KSDA have been used to identify and recognize blended expressions [1,2,9].

3. Proposed Method

The process to automate this requirement is to make the system capable of capturing images, identify the expression, identify our mood and finally play the music as per our mood. The movements and resulting positions of the facial muscles underneath the skin result in a particular facial expression relating to a unique emotion. This kind of facial expressions are noticed even in some mammals and animal species. The social interaction between human beings include messages conveyed through these face expressions.

3.1 Description and Algorithm

Mathematically, the Action Units and the centroid information can be used to describe the face expression. When the related concepts of physics and mathematics behind capturing of images are explored, it is observed that the face expressions can be divided with their respective aspect ratios and centroid of the 2D picture captured by the system/webcam of a 3D human being can be calculated. For this, it is seen that the process of expression identification and automatic playing of audio will undergo typical operations as shown in Fig. 1. below.

3.2 Extraction of face expression

Accordingly, the face is recognized with aspect ratio i.e., the ratio includes height and width of the face. Based on this concept of Aspect Ratio, major expressions can be divided into their respective moods.

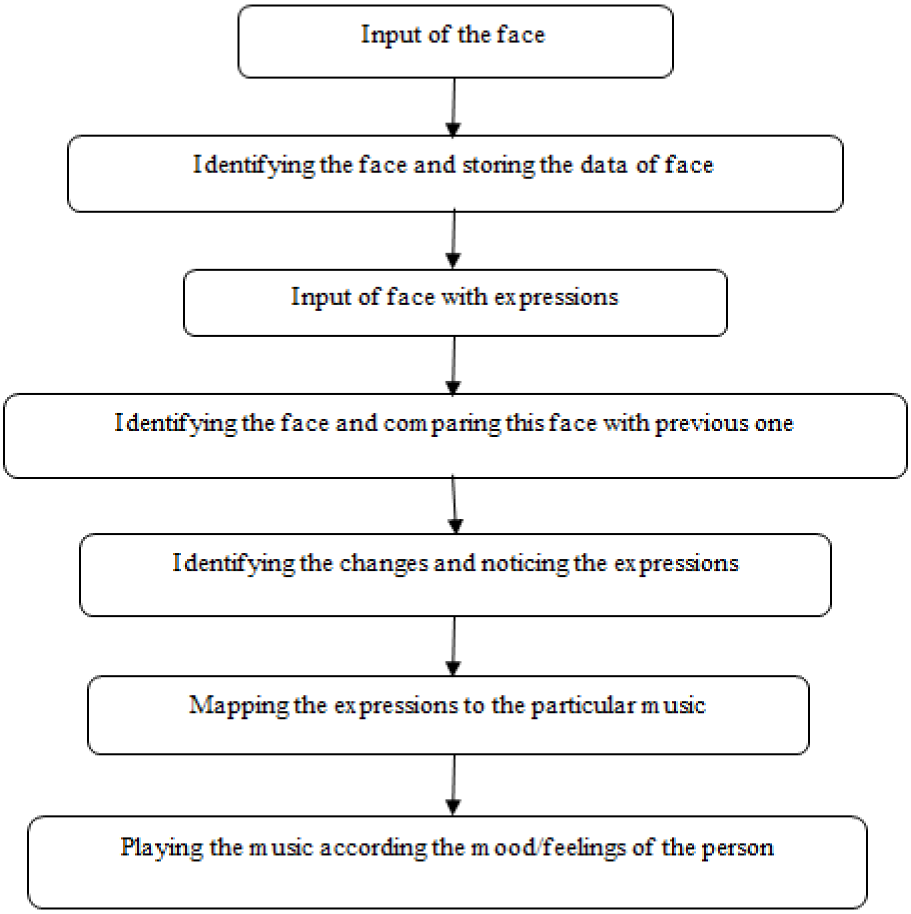

Fig. 1. Flowchart of the automation of mood-based music.

3.3 Identifying the mood or feeling of the person through Zernike Moment

The listener's state of mind is determined as per the primary emotions, as classified by Plutchik's theory of emotion [6]. In this work, the Pattern vector is defined by the features of height of the face, width of the face, the Aspect Ratio and moment generated by Zernike polynomials [7]. The facial features can be uniquely extracted through the use of only the Aspect Ratio or the Aspect Ratio in combination with the Zernike Moment, but this will result in the features for the differently distanced face to be the same. In order to ensure a more effective, unique and correct emotion classification, the height and the width of the face are also considered.

Zernike polynomials are commonly used as basis functions of image moments. These polynomials are orthogonal by nature to each other and hence, can be used to represent the image properties uniquely. These polynomials are invariant to rotation, scale and translation. Few data points are sufficient to determine the descriptors accurately [7]. Once the features are extracted, their combination is identified to be a corresponding expression, as given in the example.

3.4 Song selection according to the mood

The song is mapped with the expression on the person's face through the use of a play-list record. Finally, the selected music or song is played automatically.

4. Implementation

As an initial step, the range of the centroids are mapped to the person's mood and stored as a file programmatically. This file is used as a reference. Twelve moods have been captured for a particular subject in hand.

The individual's picture is captured through the webcam on an I5 quad core processor, with a configuration of 4GB RAM, 1Tb HDD. The face is detected and consecutively extracted. The features of the face required to identify the expression are obtained through Pattern Vector generation. The corresponding emotion is identified through the vector value. The code was first implemented using the Corpus images shown in Fig. 2. (a) to test for understanding the working of the algorithm.

Fig. 2. (a) Corpus images showing twelve types of emotions [6].

The second step involving the location identification and extraction of the faces from the pictures is seen in Fig. 2.(b). Observation of different expressions simultaneously on the upper and the lower face represent different emotions, called blended facial expressions of emotions. An example of these expressions with the Action Units (AUs) is shown in Fig. 3.

Fig. 2. (b) Corpus images showing twelve types of emotions [6].

A set of 12 blended expressions representing the emotions of tranquility, enjoyment, amazement, soulfulness, crying, normal, hard feelings, unbelieving, smiling, hysterical emotion and questioning have been considered for this study. For this special coordinate system of the face, the x- and y- axes are indicated by the line joining the inner corners of eyes and its perpendicular line respectively.

A U6
A U12
A U20

AU2
AU10
AU15
AU6

AU2
AU10
AU15

Fig. 3. An example representing the Action Units relating to blended facial expression (a) A Happy expression; (b) A Surprised Expression; (c) A Happily Surprised Expression.

Upper and lower facial features are defined by fifteen and nine parameters for ease of understanding [3]. The ratio of the current feature value (i.e., expressive one) to the neutral feature is used to normalize the face feature parameters. Two such samples of these normalized parameters are listed in Table 1. below.

Table 1. Example of Mathematical Notations for the normalized parameters

Normalized Parameter	Mathematical Representation
Inner brow motion	$\bar{a}_{bi}$
Outer brow motion	$\bar{a}_{bo}$

It is assumed that face expressions can be represented approximately as combinations of several Action Units (AUs) which are represented in terms of normalized parameters. Assuming AU1 symbolizes inner brow rise i.e., inner brow motion and positive y-axis direction $\bar{a}_{bi}$ indicates large feature moment with respect to neutral one in y direction so AU1 - $\bar{a}_{bi}$ .The facial expression, its AU combination and its symbolic representation is given for three example expressions in Table 2 below:

Table 2. List of three examples for emotion representation of Facial Expressions [3]

Facial Expression	Corresponding Action Unit	Proposed Symbolic Representation
Happy	AU12, AU25 (primary), AU6	$\bar{a}_{bi}\mathbf{x} + \bar{a}_{lh}\mathbf{y} + \bar{a}_{che}\mathbf{y}$
Sad	AU4, AU15 (primary), AU1, AU6, AU17	$-[\bar{a}_{bi}\mathbf{y} + \bar{a}_{bo}\mathbf{y} + \bar{Q}_{b}\mathbf{x}]-[\bar{a}_{ll}\mathbf{y} + \bar{a}_{lr}\mathbf{y}] + \bar{a}_{che}\mathbf{y} + \bar{a}_{che}\mathbf{y}$
Surprised	AU1, AU2, AU25, AU26 (primary)	$\bar{a}_{bi}\mathbf{y} + \bar{a}_{bo}\mathbf{y} + 2\bar{a}_{lh}\mathbf{y}$

The table indicating the corresponding Zernike values for six faces is listed in the Table 3, with the initial value being taken as 0.

Table 3. The range of the centroid relevant to the facial expression for the Corpus Image

Face Number	Centroids	Aspect Ration	Emotion
face5.bmp	52012		tranquil
face3.bmp	54571	2558.44	peaceful
face1.bmp	55046	475.26	enjoying
face6.bmp	59748	4701.88	amazed
face4.bmp	62618	2869.89	soulful
face8.bmp	65914	3296.18	crying

The songs obtained for two of the emotions are given in Fig. 4.

/Desktop/res/face11.bmp
aspectratio new:0.85106385
zernike new: 61558.753690701145
song: Desktop/programs/gesturereco/songs/soulful.wav

/Desktop/res/face4.bmp
aspectratio new:0.8636364
zernike new: 54032.30394614844
song: /Desktop/programs/gesturereco/songs/peaceful.wav

Fig. 4. The file links to two sample emotions.

This was implemented on a few images from the Cohn Kanade standard database as a sample. Once the algorithm was proved successful, it was then implemented on live data. The emotions of the listener will not change abruptly, but will transform gradually, after a 3 to 5 second period. To identify the emotion in the listener, an image is captured within 5 seconds approximately at the rate of one image per second. The facial expression is identified within a period of 30 seconds and the song associated with the emotion is played. Two sample results that were obtained with this implementation are shown in Fig. 5. below indicating the process of capturing the image, face extraction, mapping of the centroid to a particular emotion and then obtaining the corresponding audio in the form of a song. It is observed that when Zernike moments are considered as feature extractors, the noise from the face images has less effect on the classification of emotions. This was implemented by recording twelve emotions of ten people to run the test.

Fig. 5. (a) Process to represent sad mood

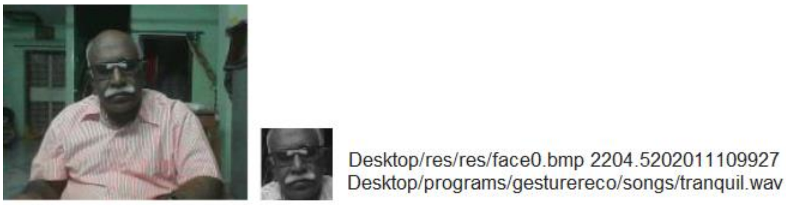

Fig. 5. (b) Process to represent tranquil mood

5. Limitations

The process was implemented for twelve categories of emotions only. However, additional moods whose centroid values lie in the range are identified with a reasonable range of the centroid values, instead of only distinct values. The complex emotions of a person cannot be detected in a single image. Also, proper illumination of the room is required for the correct recognition of the facial expression. The centroids for the facial expressions vary from one individual to the other. This requires identifying the range for the individual initially.

6. Future Work

More work needs to be conducted to make the process illumination-invariant. The emotions and their respective range of centroid values can be extended to recognize any other mood. Further study on this aspect is required based on the Pattern Vector used in this algorithm. An automatic training of the system to generate the respective ranges would make the process more versatile.

7. Conclusion

This paper focuses on the mathematical formulation of the expressions related to the face, representing different emotions. A total of twelve of the basic human facial expressions are identified based on these formulations and used for the application. The proposed method is implemented with a few example expressions.

Acknowledgments. The author is thankful to the support team for the implementation of this project within the prescribed time and utilization of the resources in the lab

References

[1] Dubey, M., and Singh, L.: Automatic Emotion Recognition Using Facial Expression: A Review, International Research Journal of Engineering and Technology, vol. III, no. 02 (2016) pp. 488-492

[2] Happy, S.L., Dasgupta, A., Patnaik, P., and Routray, A.: Automated Alertness and Emotion Detection for Empathic Feedback during e-Learning, in IEEE Fifth International Conference on Technology for Education, Kharagpur (2013)

[3] Saha, P., Bhattacharjee, D., De, B. K. , Nasipuri, M.: Mathematical Representations of Blended Facial Expressions towards Facial Expression Modeling, in 7th International conference on Intelligent Human Computer Interaction (IHCI), Procedia Computer Science 84 (2016) pp. 94 – 98

[4] Du, S., Tao, Y., Martinez, A. M.: Compound facial expressions of emotion, Proc. of the National Academy of Sciences of USA (2014) 111: E1454– E1462

[5] Kezia, S., Prabha, I. S., Kumar, V. V.: A color-texture based segmentation method to extract object from background, International Journal of Image, Graphics and Signal Processing 3 (2013) pp. 19-25

[6] Abbasi, M. M., Beltiukov, A.: Summarizing Emotions from Text Using Plutchik's Wheel of Emotions, Scientific Conference on Information Technologies for Intelligent Decision making Support, vol. 166 (2019) pp. 291-294

[7] Khotanzad, A., Hong, Y. H.: Invariant Image Recognition by Zernike Moments, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 12, No. 5, (1990) pp. 489-497

[8] Thaware, R.: Real-Time Face Detection and Recognition with SVM and HOG Features, Embedded Systems (2018): https://www.eeweb.com/

[9] Soladie, C., Stoiber, N, Seguier, R.: A New Invariant Representation of facial expressions: Definition and Application to Blended Expression Recognition, 19th IEEE International Conference on Image Processing (2012) DOI: 10.1109/ICIP.2012.6467435