International Journal of
Scientific Engineering and Technology Research

ISSN 2319-8885
Vol.07,Issue.05,
May-2018,
Pages:0979-0982

www.ijsetr.com

Music Therapy using Face Recognition

SHAHLI[1], ARUN T NAIR[2]

[1]PG Scholar, Dept of ECE, KMCT College of Engineering And Technology, Kallanthode, Calicut, India,
E-mail: saheersshahlisaheer@gmail.com.

[2]Assistant Professor, Dept of ECE, KMCT College of Engineering And Technology, Kallanthode, Calicut, India,
E-mail: ckarunlal@gmail.com.

Abstract: In present day technology human-machine interaction is growing in demand and machine needs to understand human gestures and emotions. If a machine can identify human emotions, it can understand human behavior better, thus improving the task efficiency. Emotion recognition from facial images is a very active re-search topic in human computer interaction (HCI) Emotions can understand by text, vocal, verbal and facial expressions. Facial expressions play big role in judging emotions of a person. In this project, we propose a method for real time emotion recognition from facial image. In the proposed method we use three steps face detection using Haar cascade, features extraction using Active shape Model(ASM), (26 facial points extracted) and Multi SVM classifier for classification of five emotions anger, disgust, happiness, neutral and sad. The proposed method is implemented in MATLAB, emotion recognition done using the trained dataset and an average accuracy of 94% is achieved.

Keywords: Face Detection, LBP, KW Feature Selection, CK Database.

I. INTRODUCTION

Due to the ongoing growth along with the extensive use of smart phones, services and applications, emotion recognition is becoming an essential part of providing emotional care to people. Provisioning emotional care can greatly enhance users experience to improve the quality of life. The conventional method of emotion recognition may not cater to the need of mobile application users for their value-added emergent services. Moreover, because of the dynamicity and heterogeneity of mobile applications and services, it is a challenge to provide an emotion recognition system that can collect, analyze, and process emotional communications in real time and highly accurate manner with a minimal computation time. There exist a number of emotion recognition systems in the literature. The emotion can be recognized from speech, image, video, or text. There are many applications of the emotion recognition in mobile platforms. In mobile applications, for example, the text of the SMS can be analyzed to detect the mood or the emotion of the users. Once the emotion is detected, the system can automatically put a corresponding `emoji' in the SMS. By analyzing a video in the context of emotion, a smart phone can automatically change the wallpaper, or play some favorite songs to coop with the emotion of the user. The same can be applied using oral conversation through a smart phone; an emotion can be detected from the conversational speech, and an appropriate filtering can be applied to the speech. To realize an emotion-aware mobile application, the emotion recognition engine must be in real-time, should be computationally less expensive, and give high recognition accuracy [1].

Most of the available emotion recognition systems do not address all of these issues, as they are designed mainly to work offline and for desktop applications. These applications are not bounded by storage, processing power, or time. For example, many online game applications gather the video or the sound of the users, process them in a Cloud, and analyze them for a later improvement (next version) of the game [6]. In this case, the Cloud can provide unlimited storage and processing power, and the game developer can have enough time to analyze. On the contrary, emotion aware mobile applications cannot have this luxury. Mehmood and Lee proposed an emotion recognition system from brain signal pattern using late positive potential features. This system needs EEG sensors to be associated to the mobile device, which is an extra burden to the device. Chen et al. proposed CP-Robot for emotion sensing and interaction. This proposed robot uses a Cloud for image and video processing. A distant learning system LIVES through interactive video and emotion detection. The LIVES can recognize the emotion of the students from the video, and adjust the content of the lecture. In this paper, we propose a high-performance emotion recognition system for mobile applications. The embedded camera of a smart phone captures video of the user. The Bandlet transform is applied to some selective frames, which are extracted from the video, to give some subband images. Local binary patterns (LBP) histogram is calculated from the subband images. This histogram describes the features of the frames. A Gaussian mixture model (GMM) based classier is used as a classier. The proposed emotion recognition system is evaluated using several databases. The contribution of this paper is as follows: (i) the use of the Bandlettrans form inemotion recognition, (ii) the use of the Kruskal-Wallis (KW) feature selection to reduce the time requirement during a test phase, and (iii) an achievement of higher accuracies in two publicly available databases compared with those using other contemporary systems.

II. RELATED WORK

A. Proposed Emotion Recognition System

Fig1 shows a block diagram of the proposed emotion recognition system for mobile applications. In the following subsections, we describe the components of the proposed system.

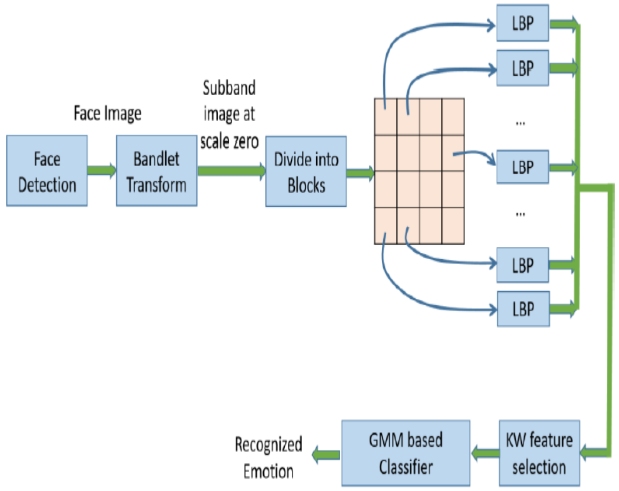


Fig.1. Block diagram of the proposed emotion recognition system.

III. IMPLEMENTATION

A. Capturing Video

An embedded camera of the smart phone captures the video of the user. As most of the time, the user faces towards the screen of the phone, the video mainly captures the face, the head, and some body parts of the user.

B. Selecting Representative Frames

As there are many frames in the video sequence, we need to select some representative frames from the sequence to reduce the burden of the processing. To select frames, first, all the frames are converted to the gray scale. Then, histograms are obtained from each frame. A chi-square distance is calculated between the histograms of two successive frames. We select a frame, when the distances between the histogram of this frame and that of the previous frame, and the next frame are minimal. In this way, we select a frame, which is stable in nature.

C. Face Detection

Once we select the frames, the face areas in the frames are detected by the Viola-Jones algorithm. This algorithm works fast, and is suitable for a real-time implementation. Now a day, many smart phones have the face detection functionality embedded into the mobile system.

D. Bandlet Transform

The Bandlet transform is applied to the detected face area. A face image has many geometric structures that carry valuable information about the identity, the gender, the age, and the emotion of the face. A traditional wavelet transform does not take care much about the geometric structure of an image, especially in sharp transitions; however, representing sharp transitions using geometrical structures can improve the representation of image. One of the major obstacles of using geometrical structure is a high computation complexity. The Bandlet transform overcomes the obstacle to represent geometric structure of an image by calculating the geometric flow in the form of Bandlet bases. This transform works on the gray scale images; therefore, in the proposed method, the input color images are converted into gray scale images. To form orthogonal Bandlet bases, the image needs to be divided into regions consisting of geometric flows. In fact, the image is divided into small square blocks, where each block contains at most one contour. If a block does not contain a contour, the geometric flow in that block is not defined. The Bandlet transform approximates the regions (Ω) by using the wavelet basis in the L2 (Ω) domain as follows.

$$\begin{Bmatrix} \varphi_{i,n}(x) = \varphi_{i,n1}(x1)\varphi_{i,n2}(x2) \\ \psi^{H}_{i,n}(x) = \varphi_{i,n1}(x1)\psi_{i,n2}(x2) \\ \psi^{V}_{i,n}(x) = \psi_{i,n1}(x1)\varphi_{i,n2}(x2) \\ \psi^{D}_{i,n}(x) = \psi_{i,n1}(x1)\psi_{i,n2}(x2) \end{Bmatrix} \quad (1)$$

Where, ψ (.) and φ(.) are the wavelet and the scaling functions, i is the dilation, n1 xn2 is the dimension of the input image, x is the pixel location. $\varphi, \psi^{H}, \psi^{V}, \psi^{D}$ are the coarse level (low-frequency) approximation, high-frequency representations along horizontal, vertical, and diagonal directions of the image. Fig.2 illustrates the Bandlet decomposition.

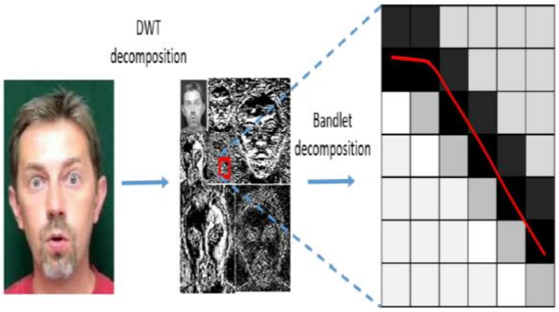


Fig.2. Illustration ofTheBandlet Decomposition.

Once the wavelet bases are computed, the geometric flow is calculated in the region by replacing the wavelet bases by the Bandlet orthonormal bases as follows.

$$\begin{Bmatrix} \varphi_{i,n}(x) = \varphi_{i,n1}(x1)\varphi_{i,n2}(x2 - c(x1)) \\ \psi^{H}_{i,n}(x) = \varphi_{i,n1}(x1)\psi_{i,n2}(x2 - c(x1)) \\ \psi^{V}_{i,n}(x) = \psi_{i,n1}(x1)\varphi_{i,n2}(x2 - c(x1)) \\ \psi^{D}_{i,n}(x) = \psi_{i,n1}(x1)\psi_{i,n2}(x2 - c(x1)) \end{Bmatrix} \quad (2)$$

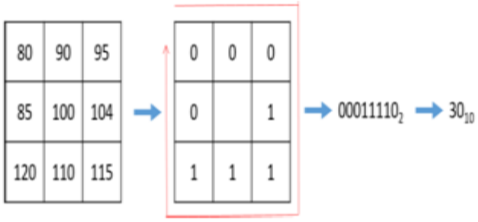
In the above equation, c(x) is the geometric flow line of the fix translation parameter x1 as follows.

$$c(x) = \sum_{u=x_{min}}^{x} c^{'(u)} \tag{3}$$

From the above equation, we understand that the block sizes affect the geometric flow direction. In practice, the smaller block size gives better representation of geometric flow than the larger block size. In our proposed system, we investigated the effect of different block sizes, and scale decompositions on the emotion recognition accuracy.

E. LBP

The next step of the proposed system is to apply the LBP on the subband images of the Bandlet transform. The LBP is a powerful gray-level image descriptor, which operates in real-time. It has been used in many image-processing applications including face recognition, gender recognition, and ethnicity recognition. The basic LBP operates in a 3 x 3neighborhood, where the neighboring pixels' intensities are threshold by the center pixel intensity. If a pixel intensity is higher than the center pixel's intensity, then a `1' is assigned to that pixel position, otherwise a `0' is assigned. All the eight neighboring pixels' assigned values are concatenated to produce an 8-bit binary number, which is next converted to a decimal value. This decimal value corresponds to the LBP value of the center pixel. A histogram is created from these LBP values to describe the image. For an interpolated LBP, a circular interpolation of radius R and P number of pixels is used. A uniform LBP is a pattern, where there are at most two bit-wise transitions from zero to one or vice versa. Fig.3 illustrates the LBP calculation.



80	90	95
85	100	104
120	110	115

0	0	0
0		1
1	1	1

$\rightarrow 00011110_2 \rightarrow 30_{10}$

Fig.3.Illustration of the LBP calculation.

In the proposed system, the subband obtained afterthe Bandlet decomposition is divided into non-overlapping blocks. The LBP histograms are calculated for all the blocks and then concatenated to describe the feature set for the image.

F. KW Feature Selection

The number of features for the image is very huge; many features slow down the process, and they also bring the `curse of dimensionality'. Therefore, a feature selection technique can be applied to reduce the dimension of the feature vector. There are many feature selection techniques, each of which has its own advantage and disadvantage. In our proposed system, we adopt the KW technique for its simplicity and low computational complexity, keeping in mind that the system is to deploy in mobile applications. The KW is a non-parametric one-way analysis of variance test, where the null hypothesis is that the samples from different groups have the same median. It returns a value p; if the value of p is close to zero, the null hypothesis is rejected, and the corresponding feature is selected. The feature, which results in a big value of p, is discarded because it is considered to be non-discriminative.

G. GMM-Based Classifier

The selected features are fed into a GMM based classifier. During training, models of different emotions are created from the feature set. During testing, log-likelihood scores are obtained for each emotion using the feature set and the models of the emotion. The emotion corresponding to the maximum score is the output of the system. In the experiments, different numbers of Gaussian mixtures are investigated. We choose the GMM based classifier because it can operate in real-time, it is more stable than the neural network based classifiers

IV. EXPERIMENTAL RESULTS AND ANALYSIS

To validate the proposed system, we used a number of experiments using two publicly available databases, namely Canade-Kohn (CK) and the Japanese female facial expression (JAFFE) databases. In the following subsections, we briefly describe the databases, and present the experimental results and discussion.

V. DATABASES

In our experiments, we used two databases. The JAFFE database consists of emotional face images of Japanese actresses. There are total 213 face images of 10 female Japanese. All the images are gray, and have a resolution of 256 x 256. The original images were printed, scanned, and digitized. The faces are frontal. There are seven emotion classes, which are anger, happiness, sadness, disgust, afraid, and surprise, in addition to neural. The CK database was created by the faces of 100 university-level students. After careful observations, the faces of four students were discarded because they were not properly showing the emotions. The students were ethnically diverse. Video sequences were captured in a controlled environment, and the participants posed with six emotions as mentioned previously. There were 408 video sequences, and three most representative image frames were selected from each sequence. The first frame of each sequence was treated as a neutral expression. The total number of images is 1632.

VI. EXPERIMENTAL RESULTS AND DISCUSSIONS

First, we used the JAFFE database in our experiments to set up different parameters of the proposed system. We chose this because it a smaller size database than the CK database. Once we fix the parameters using the JAFFE database, we used the CK database. During classification, we adopted a 5-fold approach, where the database was divided into five equal groups. In each iteration, four groups were trained and the rest was tested. After five iterations, all the five groups were tested. With this approach, the biasness of the system towards any particular images could be reduced.

For feature selection using the KW method, we set the value of p to be 0.2, which gave the optimal results for almost all the case as shown in Figs.4 to 7.

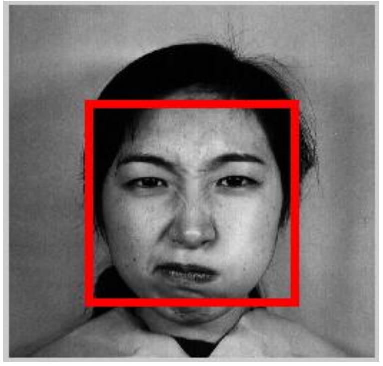


Fig.4. Viola Jones Face Detection.

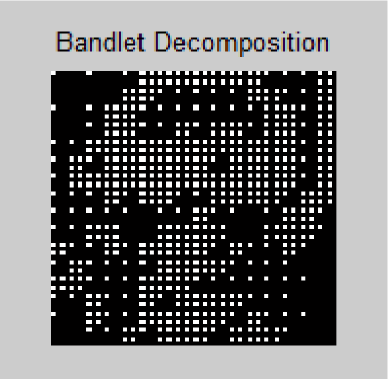


Fig.5.Bandlet decomposition.

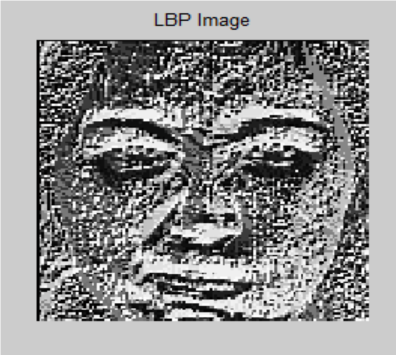
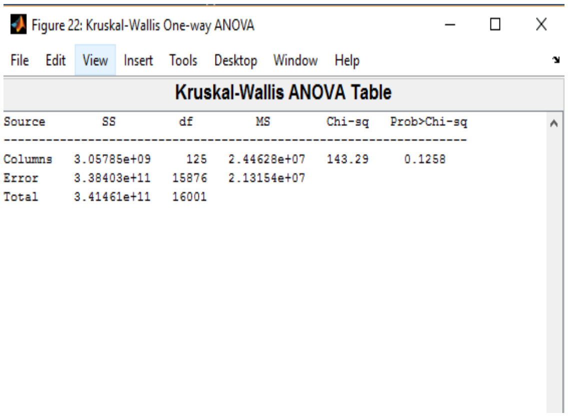


Fig.6.LBP image.



```
Kruskal-Wallis ANOVA Table
Source         SS          df          MS         Chi-sq   Prob>Chi-sq
----------------------------------------------------------------------
Columns    3.05785e+09      125   2.44628e+07    143.29      0.1258
Error      3.38403e+11    15876   2.13154e+07
Total      3.41461e+11    16001
```

Fig.7. KW anova feature table.

VII. CONCLUSION

An emotion recognition system for mobile applications has been proposed. The emotions are recognized by face images. The Bandlet transform and the LBP are used as features, which are then selected by the KW feature selection method. The GMM based classifier is applied to recognize the emotions. Two publicly available databases are used to validate the system. The proposed system achieved 99.8% accuracy using the JAFFE database, and 99.7% accuracy using the CK database. It takes less than 1.4 seconds to recognize one instance of emotion. The high performance and the less time requirement of the system make it suitable to any emotionaware mobile applications. In a future study, we want to extend this work to incorporate different input modalities of emotion.

VIII. REFERENCES

[1] M. ShamimHossain, Ghulam Muhammad “An Emotion Recognition System for Mobile Applications” IEEE Access, Received January 30, 2017, accepted February 8, 2017

[2] M. S. Hossain and G. Muhammad, ``Audio-visual emotion recognition using multi-directional regression and Ridgelet transform," J. Multimodal User Interfaces, vol. 10, no. 4, pp. 325_333, Dec. 2016.

[3] M. S. Hossain, G. Muhammad, M. F. Alhamid, B. Song, and K. Al-Mutib, ``Audio-visual emotion recognition using big data towards 5G," Mobile Netw. Appl., vol. 21, no. 5, pp. 753_763, Oct. 2016.

[4] M. S. Hossain, G. Muhammad, B. Song, M. M. Hassan, A. Alelaiwi, and A. Alamri, ``Audio_visual emotion-aware cloud gaming framework," IEEE Trans. Circuits Syst. Video Technol., vol. 25, no. 12, pp. 2105_2118, Dec. 2015

[5] J. C. Castillo et al., ``Software architecture for smart emotion recognition and regulation of the ageing adult," Cognit. Comput., vol. 8, no. 2, pp. 357_367, Apr. 2016.

[6] M. Chen, Y. Zhang, Y. Li, M. M. Hassan, and A. Alamri, ``AIWAC: Affective interaction through wearable computing and cloud technology," IEEE Wireless Commun., vol. 22, no. 1, pp. 20_27, Feb. 2015.

[7] C. Shan, S. Gong, and P. W. McOwan, ``Facial expression recognition based on local binary patterns: A comprehensive study," Image Vis. Comput., vol. 27, no. 6, pp. 803_816, 2009.

[8] K. K. Rachuri, M. Musolesi, C. Mascolo, P. J. Rentfrow, C. Longworth, and A. Aucinas, ``EmotionSense: A mobile phones based adaptive platform for experimental social psychology research," in Proc. 12th ACM Int. Conf. Ubiquitous Comput. (UbiComp), Copenhagen, Denmark, Sep. 2010, pp. 281_290.

[9] Y. Zhang, M. Chen, S. Mao, L. Hu, and V. Leung, ``CAP: Community activity prediction based on big data analysis," IEEE Netw., vol. 28, no. 4, pp. 52_57, Jul./Aug. 2014.

[10] W. Zhang, D. Zhao, X. Chen, and Y. Zhang, ``Deep learning based emotion recognition from Chinese speech," in Proc. 14th Int.Conf. Inclusive Smart Cities Digit.Health (ICOST), vol. 9677. New York, NY, USA, 2016, pp. 49_58.