Automatic Facial Expression Analysis of Students in Teaching Environments

Chuangao Tang, Pengfei Xu(✉), Zuying Luo(✉), Guoxing Zhao, and Tian Zou

Beijing Key Laboratory of Digital Preservation and Virtual Reality for Cultural
Heritage, College of Information Science and Technology,
Beijing Normal University, Beijing 100875, China
{xupf,luozy}@bnu.edu.cn

Abstract. Based on students' facial expressions, the teacher in class can know the students' comprehension of the lecture, which has been a standard of teaching effect evaluation. In order to solve the problem of high cost and low efficiency caused by employing human analysts to observe classroom teaching effect, in this paper we present a novel and high-efficiency prototype system, that automatically analyzes students' expressions. The fusion feature called Uniform Local Gabor Binary Pattern Histogram Sequence (ULGBPHS) is employed in the system. Using K-nearest neighbor (KNN) classifier, we obtain an average recognition rate of 79% on students' expressions database with five types of expressions. The experiment shows that the proposed system is feasible, and is able to improve the efficiency of teaching evaluation.

Keywords: Teaching effect evaluation · Facial expression recognition · ULGBPHS · Feature fusion

1 Introduction

Classroom teaching evaluation has been a hot spot in recent years. It has been widely used to promote the classroom teaching quality and teachers' skills [1]. Students' grasp of the lecture and students' emotional involvement, both of which are indexes of the teaching effect evaluation, are correlated with their facial expressions. For instance, students usually smile after comprehending the lecture and finding it interesting, and they form negative expressions when they find that the content is too abstruse to understand. The research [2] indicated that facial expression ranks top of the mode of nonverbal communication, followed by body language, gestures and hands. Experienced instructors often adjust their teaching according to students' expressions during the lectures. Currently, the classical methods of instructional measurement like tests, exams, questionnaires, interviews, observations are applied in classroom teaching evaluation. It is a common phenomenon that dozens of professors are invited every semester to form a supervision team to observe the classroom sessions of required and elective courses. After the observation, the feedback reports with the ratings and comments are written by the office and provided to the corresponding faculties within a few weeks. However, these manual evaluation methods often come with high cost and low efficiency [1, 2].

J. Yang et al. (Eds.): CCBR 2015, LNCS 9428, pp. 439–447, 2015.
DOI: 10.1007/978-3-319-25417-3_52

As our research motivation is to realize evaluating classroom teaching effect automatically, we conduct an exploratory research based on the recognition results of students' facial expressions in class. Since the late 1990s, an increasing number of efforts toward automatic affect recognition were reported in literature [3]. AFER technology has been successfully applied in the areas such as human-computer interaction (HCI), driver fatigue detection, e-learning, etc. [3]. Most of the existing facial expression recognition approaches are based on posed expression databases, like Japanese Female Facial Expression (JAFFE) database, Cohn-Kanade (CK) database. Due to lack of facial expression database in pedagogical environments, the researchers in our college firstly built the students' spontaneous expressions database in classroom teaching environments. Then, we propose a prototype system to automatically analyze students' expressions in class. In this system, a fusion feature ULGBPHS is employed. K-nearest neighbor (KNN) method based on Euclidean distance is employed for classification. The proposed method obtains a high recognition rate of 96.7% on JAFFE database, which outperforms some existing methods, and a recognition rate of 79% on self-build database. Furthermore, to the best of our knowledge, using the recognition results of students' facial expressions in class for traditional classroom teaching effect evaluation is first proposed in this paper.

The rest of this paper is organized as follows. Section 2 reviews the related work. Section 3 describes the self-build students' expressions database and an overview of the proposed system is given in section 4. Details of the proposed system are presented in section 5, followed by the experimental results in section 6. In the last section we conclude this paper with discussion.

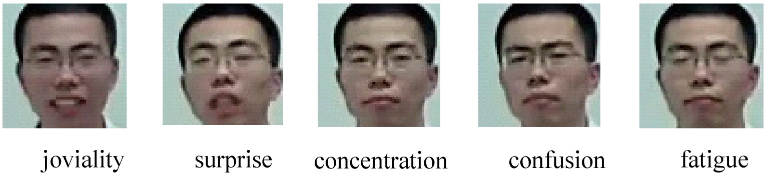
2 Related Work

In this section, we offer an overview of some recent literature about facial expression recognition (FER) and some FER systems used in learning environments. Because of the importance of face in emotion expression and perception [3], extracting an efficient representation of the face is a key step for successful FER [4]. There are two main streams in the current research of extracting features for facial expression recognition: geometric based methods and appearance based methods [4]. Geometric features contain information about the location and shape of facial features. In general, a shape model defined by 58 facial landmarks is used during the process of geometric feature extraction, in which noise and tracking errors often decline the recognition performance. Appearance based features examine the appearance change of the face (including wrinkles, bulges and furrows) and are extracted by image filters applied to the face or sub regions of the face [3, 4]. Appearance based features are less reliant on initialization and can encode micro patterns in skin texture that are important for facial expression recognition. Gabor feature [6] and extended LBP feature [5] are widely used as appearance based features in facial expression recognition approaches. Hua Lu et al. [5] presented a method of divided local binary pattern (DLBP) and obtained a recognition rate of 95.7% on JAFFE database. Seung Ho Lee et al. [6] proposed a new sparse representation based FER method and got the highest overall recognition rate of 94.7% on JAFFE database under their experimental scenarios, compared with SRC+LBP and SRC+ Gabor, where the recognition rate of the former is 90.30% with the latter 91.21%. Recently, deep learning technology has attracted many researchers' interest. Ping Liu et al. [7] presented a novel Boosted Deep Belief Network (BDBN) framework for facial expression recognition, and obtained impressive recognition results. In their study, it took about 8 days to complete the overall training for 6 expressions in an 8-fold experimental setup on a 6-core 2.4GHz PC using Matlab implementation.

Some researchers have been focusing on facial expression and facial affect in the lab or wild [8]. In the lab environment, Whitehill et al. [9] used Gabor features with a SVM classifier to detect engagement as students interacted with cognitive skills training software. Labels used in their study were obtained from retrospective annotation of videos by human judges. While in the wild environment, Nigel Bosch et al. [8] collected the data of students' facial expressions, including videos containing students' faces and affect labels in the real-world environment of a school computer lab, where the students were interacting with a game-based physics education environment called Physics Playground.

3 Students' Spontaneous Expressions Database

Having enough labeled data of facial expressions is a prerequisite in designing automatic facial expression recognition system [3]. The self-build facial expression database in this paper contains the students' spontaneous expressions [18], as opposed to posed expressions in current mainstream databases. Since we focus on the spontaneous facial behavior correlated to learning, we predefine the labels of expression as follows: joviality, surprise, concentration, confusion, fatigue. The corresponding face images are demonstrated in Fig.1.



joviality surprise concentration confusion fatigue

Fig. 1. Five types of facial expressions

As [3] pointed out, current techniques for detection and tracking of facial expressions are sensitive to head pose, clutter, and variations in lighting conditions. Thus the experiment of self-build expression database was conducted under controlled condition to get rid of some above mentioned variations, with 23 youthful college students from different majors invited to participate in the experiment. There were 17 participants wearing glasses. They received a short-term training that the behaviors like extreme pose or position, occlusions from hand-to-face gestures, and rapid movements should be avoided during the process of experiment. All the participants sat on the seats in a common classroom, where the light condition was normal. The task for them was to watch 6 short videos, which lasted about 15 minutes. A 1080p HD camera was used for capturing their facial expressions. Another task for participants was to label their own expressions, respectively.

4 System Overview

In this paper, we propose a prototype system of AFER to analyze students' expressions for classroom teaching effect evaluation. The system consists of 5 modules: data acquisition module, face detection module, face recognition module, facial expression recognition module and post-processing module. The schematic diagram of the system is demonstrated in Fig. 2.

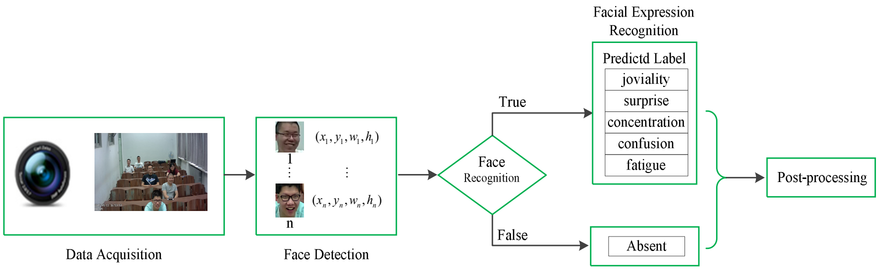


Fig. 2. Schematic diagram of proposed facial expression analysis system

Data Acquisition: A full 1080p HD camera is configured at the front of the classroom. Using HD camera can ensure every student's face enough resolution.

Face Detection: We use AdaBoost method to detect faces. The face images are segmented based on the location and size. We use canthi positions returned by the Structured Output Support Vector Machines (SO-SVM) [10] method to calculate the angle of tilt, which is used to normalize face rotation. Finally, the region of interest is cropped with geometry rules of face to remove background. We resize the cropped image region to 66 × 66 pixels.

Face Recognition: The location and size of the face are used to speed up computation if the face is recognized in the current frame, instead of repeating face recognition for the same student in the next frame. Here, we use location and size feature, known in advance, of the faces to sort the detected faces for facial expression recognition. If and only if the face is verified, facial expression recognition will be conducted. Otherwise, our system outputs an 'Absent' label.

FER: In this paper, the fusion feature ULGBPHS outperforms onefold feature. The details of the facial expression recognition will be illustrated in section 5.

Post-processing: In this stage, the results of facial expression recognition are used to evaluate classroom teaching effect, which will be analyzed in section 6.

5 Facial Expression Recognition

In this section, two main parts of our expression recognition system, feature extraction and expression classifier, are described. Feature extraction aims to build derived values, which is informative, non-redundant, and facilitates the subsequent learning steps. Expression classifier identifies which of a set of classes a new observation belongs to, on the basis of a training set containing observations whose class is known. In this system, we employ fused feature based on Gabor [11, 13] and LBP [12, 14].

Gabor feature: Gabor features have been widely used in many pattern analysis applications. C.J.Liu et al. [11] pointed that Gabor feature performs the best in classification of expression units. *2D* Gabor filter is a Gaussian kernel function modulated by a complex sinusoidal plane wave [13], defined as:

$$\varphi_{\Pi}(f,\theta,\gamma,\eta)=\frac{f^2}{\pi\gamma\eta}\exp(-(\alpha^2x'^2+\beta^2y'^2))\exp(j2\pi fx') \qquad . \qquad (1)$$
$$\begin{cases} x' = \ \ x\cos\theta + y\sin\theta \\ y' = -x\sin\theta + y\cos\theta \end{cases}$$

Gabor features with different orientations and scales are obtained by convolving the images with *2D* Gabor filters. Due to limitations on space, we only list the kernel function in this paper. More details for Gabor feature extraction are presented in paper [13].The filters are more prominent at expression-rich positions like eyebrows, eyes, mouth and nose.

LBP feature: LBP is firstly proposed by Ojala [14]. LBP is computational efficient and robust to rotation and light variations, and has been successfully used in many object classification and detection applications. The operator labels the pixels of an image by thresholding a 3×3 neighborhood of each pixel with the center value and considering the result as a binary number [5]. Given a pixel at (x_c, y_c), the resulting LBP can be expressed in decimal form as follows:

$$LBP_{P,R}(x,y)=\sum_{p=0}^{P-1}S(i_p-i_c)2^p, \quad S(i_p-i_c)=\begin{cases}1 & \text{if } i_p-i_c\geq 0\\ 0 & \text{if } i_p-i_c<0\end{cases} \qquad (2)$$

where i_c is the gray value of the pixel at (x_c, y_c), similarly, i_p (p=0, ..., p-1) are the gray values of P equally spaced pixels on a circle of radius R. This operator was extended to use neighborhoods of different sizes and capture dominant features at different scales. A Local Binary Pattern is called uniform, which is defined in Eq. (3), if it contains at most two bitwise transitions from 0 to 1 or 1 to 0 when the binary string is considered circular [12, 14].

$$LBP^{u2}_{P,R}(x,y)=\begin{cases} F(LBP_{P,R}(x,y)) & \text{if } U(LBP_{P,R})\leq 2,\\ & F(z)\in[0,(P-1)P+1], z\in[0,255]\\ (P-1)P+2 & \text{otherwise}\end{cases} \qquad (3)$$

where $U(LBP_{P,R})=\left|S(i_{P-1}-i_c)-S(i_0-i_c)\right|+\sum_{p=1}^{P-1}\left|S(i_p-i_c)-S(i_{p-1}-i_c)\right|$, $F(z)$ is an index function. A spatially enhanced feature histogram of the image $f_l(x, y)$ is defined as follows:

$$H_{i,j} = \sum_{x,y} I\{f_l(x,y) = i\} I\{(x,y) \in R_j\},\ i = 0,\cdots,L-1; j = 0,\cdots,m-1;$$
$$I(A) = \begin{cases} 1\ ,\ A \text{ is true} \\ 0\ ,\ A \text{ is false} \end{cases} \tag{4}$$

where L is the number of different bins produced by the LBP operator. Using uniform local binary pattern (ULBP) for a neighborhood where P=8, reduces the histogram from 256 to 59 bins (58 bins for uniform patterns and 1 bin for non-uniform patterns). Usually, images are divided into non-overlapping sub-regions $\{R_0, R_1, \ldots, R_{m-1}\}$ with the size of m. Then histogram for each sub-region is calculated and concatenated into a histogram sequence, which is Uniform Local Binary Pattern Histogram Sequence.

ULGBPHS feature: The proposed system employs a fused feature called Uniform Local Gabor Binary Pattern Histogram Sequence (ULGBPHS), which has been shown to be very robust to illumination changes and misalignment. The winner of the FERA 2011 AU detection sub-challenge adopted this architecture [15-16]. Firstly, Gabor filtering is performed on target expression image. Secondly, LBP is employed to filter the magnitudes in face regions, and the output is called Uniform Local Gabor Binary Pattern (ULGBP) image. Thirdly, the ULGBP image is partitioned into non-overlapping sub-regions. Then, histogram for each sub-region is calculated and concatenated into a histogram sequence, which is the fused feature ULGBPHS used in our system. The framework of ULGBPHS approach is demonstrated in Fig.3.

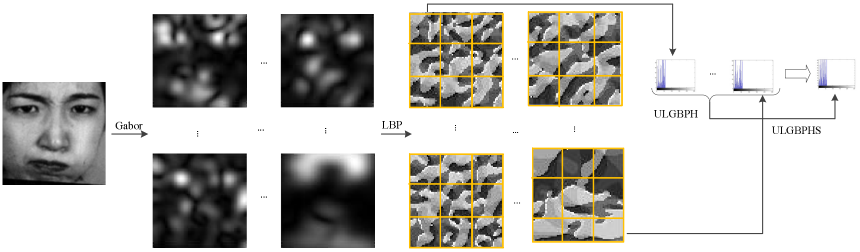


Fig. 3. The framework of ULGBPHS feature

Classification: Due to the high dimension $O(10^5)$of the fusion feature vector, the system applies two steps of dimension reduction, which are principal component analysis (PCA) and linear discriminant analysis (LDA) respectively. PCA is often considered as revealing the internal structure of the data in a way that best explains the variance in the data. LDA aims to find a linear combination of features that characterizes or separates two or more classes of data. It is proved that PCA + LDA based dimension reduction performs better than using PCA only [17]. The reduction step is defined as:

$$z_i = W_{LDA}^T \bullet W_{PCA}^T x_i \quad (i = 1, 2, \cdots, M)\ , \tag{5}$$

where x is fusion feature vector corresponding to feature extraction, z is final feature vector corresponding to feature selection, M is the number of samples. In our system, we classify the data of expression into 5 categories, thus the dimension of vector z is at most 4. Last but not the least, K-nearest neighbors algorithm (K-NN) is used as our classifier. Given a test sample, K-NN firstly finds k closest training samples in the feature space, and then uses the class membership of these k training samples to vote for the class membership of the given sample. The distance measure is often calculated based on Euclidean distance.

6 Experimental Results

Firstly, we tested our algorithm on JAFFE database. The dataset contains ten females with six types of prototypical expressions (happiness, anger, sadness, fear, surprise, disgust) and a neutral expression. There are 213 pictures with each person having 2-4 pictures of onefold expression. We selected 211 pictures which were labeled correctly. These pictures were cropped and resized to the size of 66×66 pixels for 3-fold cross-validation experiments. 2-3 pictures of each facial expression for each person were used for training and the remaining pictures were used as the test set. The intensity of each picture in the experiment was normalized. The recognition rate was obtained under person-dependent condition, as person-independent FER had not obtained satisfied results compared with person-dependent FER. The recognition results of KNN (K=3) are demonstrated in Table 1. We performed the experiment on a 4-core 3.2GHz PC with 16GB memory using Matlab implementation. As can been seen from the Table 1, the fusion feature outperforms onefold feature.

Table 1. Comparisons of different methods with corresponding recognition rate on JAFFE database. Time represents time consumption while performing feature extraction on our experimental platform, Pro. $\text{Mat}_{\text{PCA(95\%)}}$ and Pro. Mat_{LDA} are the projection matrixes for PCA and LDA, respectively. LBP operator means using only Eq.(2) method without histogram for extracting features,ULBPHS_{8x8} means that the partition grid for the image is 8x8 while extracting corresponding features, as well as ULGBPHS_{8x8}, Gabor_{5x8} means calculating Gabor filter responses at five different scales and eight different orientations.

Methods	Recognition (%)	Dimensions	Time(ms)	Pro. Mat. PCA(95%)	Pro.Mat. $_{\text{LDA}}$
LBP operator	75.2	4096	63.7	4096 × 121	121 × 6
ULBPHS_{8x8}	81.0	3776	23.8	3776 × 108	108 × 6
Gabor_{5x8}	95.2	174240	412.4	174240 × 65	65 × 6
ULGBPHS_{8x8}	96.7	151040	633.7	151040 × 115	115 × 6

Table 2. The average recognition rate of 4-fold cross-validation (K=3) on self-build expression database (%)

Methods	Expressional	Labels				
	fatigue	confusion	concentration	surprise	joviality	*Mean*
LBP operator	60.0	57.5	47.5	50.0	82.5	59.5
ULBPHS_{8x8}	47.5	67.5	65.0	65.0	87.5	66.5
Gabor_{5x8}	72.5	72.5	60.0	72.5	95.0	74.5
ULGBPHS_{8x8}	67.5	80.0	72.5	82.5	92.5	79.0

Secondly, we performed the same methods on our self-build facial expression database. Different participants have different feelings while watching the same video clip. Due to lack of expressional labels of some participants, we selected 10 qualified participants as our research objects. The participants' expressions with peak frames were selected for expression recognition. Thus, the whole dataset size is 4×5×10=200. Similarly, we used 3 pictures of each facial expression for each person to form training set, and the remaining samples for testing. The intensity of each picture in the experiment was normalized. The recognition rate is demonstrated in Table 2.

In the comparison of 5 types of expressions, joviality expression has been classified with a higher accuracy, as there is a similar laugh among different persons, which has been proven by the well- known Facial Action Coding System (FACS). While fatigue expression has been classified with the lowest accuracy, for its intensity is relatively low. Besides, compared with the other expressions, it is related to head pose as far as the participants in our experiment are concerned and it changes from person to person.

In the last stage, based on participants' expressions changing slightly within 2-3 seconds, the video frames are analyzed by the proposed system with a sampling ratio of 1:100, which also decreases the computational quantity. In this way, the system can analyze students' expressions in class in a fair short time.

7 Conclusion and Discussion

Although AFER is widely used in e-learning currently, there are few systems for analyzing students' expressions in classroom teaching environments. In this paper, we firstly explore applying the results of AFER to traditional classroom teaching effect evaluation. Compared with manual methods of analyzing teaching effect, the computer-aided facial expression analysis system improves the efficiency of evaluation substantially.

In this paper, we focus on improving the recognition rate of expression in classroom environments, and the fusion feature ULGBPHS is employed. The proposed system gets higher accuracy compared with onefold feature at the cost of increased computational consumption. In the future, we would further improve the efficiency and accuracy of this expression recognition system, and collect more data of spontaneous facial expression in the real-world environment. Since deep learning has robust performance in many machine learning applications, we will also employing this hot technology in our system, thus increase practical value of our system in instructional evaluation.

Acknowledgments. We wish to thank 23 participants and Professor Bo Sun in our university for their support. This paper is supported by the Fundamental Research Funds for the Central Universities (No. 2013YB71).

References

1. Wen, S.H., Xu, J.S., Carline, J.D., Zhong, F., Zhong, Y.J., Shen, S.J.: Effects of a teaching evaluation system: a case study. J. International Journal of Medical Education **2**, 18–23 (2011)
2. Sathik, M., Jonathan, S.G.: Effect of facial expressions on student's comprehension recognition in virtual educational environments. SpringerPlus **2**(1), 1–9 (2013)

3. Zeng, Z., Pantic, M., Roisman, G., Huang, T.S.: A survey of affect recognition methods: Audio, visual, and spontaneous expressions. IEEE Transactions on Pattern Analysis and Machine Intelligence **31**(1), 39–58 (2009)
4. Moore, S., Bowden, R.: Local binary patterns for multi-view facial expression recognition. Computer Vision and Image Understanding **115**(4), 541–558 (2011)
5. Lu, H., Yang, M., Ben, X., Zhang, P.: Divided Local Binary Pattern (DLBP) Features Description Method For Facial Expression Recognition. J Journal of Information & Computational Science **11**(07), 2425–2433 (2014)
6. Lee, S.H., Plataniotis, K., Konstantinos, N., Ro, Y.M.: Intra-class variation reduction using training expression images for sparse representation based facial expression recognition. IEEE Transactions on Affective Computing **5**(3), 340–351 (2014)
7. Liu, P., Han, S., Meng, Z., Tong, Y.: Facial expression recognition via a boosted deep belief network. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1805–1812. IEEE (2014)
8. Bosch, N., D'Mello, S., Baker, R., Ocumpaugh, J., Shute, V., Ventura, M., Wang, L., Zhao, W.: Automatic detection of learning-centered affective states in the wild. In: Proceedings of the 20th International Conference on Intelligent User Interfaces, pp. 379–388. ACM (2015)
9. Whitehill, J., Serpell, Z., Lin, Y.C., Foster, A., Movellan, J.R.: The Faces of Engagement: Automatic Recognition of Student Engagement from Facial Expressions. IEEE Transactions on Affective Computing **5**(1), 86–98 (2014)
10. Uřičář, M., Franc, V., Hlaváč, V.: Detector of facial landmarks learned by the structured output SVM. VISAPP **12**, 547–556 (2012)
11. Liu, C., Wechsler, H.: A gabor feature classifier for face recognition. In: Eighth IEEE International Conference on Computer Vision 2, pp. 270–275. IEEE (2001)
12. Chan, C.-H., Kittler, J., Messer, K.: Multi-scale Local Binary Pattern Histograms for Face Recognition. In: Lee, S.-W., Li, S.Z. (eds.) ICB 2007. LNCS, vol. 4642, pp. 809–818. Springer, Heidelberg (2007)
13. Shen, L.L., Bai, L., Fairhurst, M.: Gabor wavelets and general discriminant analysis for face identification and verification. Image and Vision Computing **25**(5), 553–563 (2007)
14. Ojala, T., Pietikäinen, M., Mäenpää, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence **24**(7), 971–987 (2002)
15. Zhang, W., Shan, S., Gao, W., Chen, X., Zhang, H.: Local gabor binary pattern histogram sequence (lgbphs): a novel non-statistical model for face representation and recognition. In: Tenth IEEE International Conference on Computer Vision 1, pp. 786–791. IEEE (2005)
16. Almaev, T.R., Valstar, M.F.: Local gabor binary patterns from three orthogonal planes for automatic facial expression recognition. In: 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction (ACII), pp. 356–361. IEEE (2013)
17. Deng, H.B., Jin, L.W., Deng, H.B., Jin, L.W.: Facial Expression Recognition Based on Local Gabor Filter Bank and PCA+ LDA. J. Journal of Image and Graphics **12**(02), 322–329 (2007)
18. BNU-LSVED Database. http://www.bnusei.net:8080/BNULSVED/cn_index.html