Classroom Teaching Feedback System Based on Emotion Detection

Sheng Chen [a,*], Jianbang Dai [b], and Yizhe Yan [c]

College of Automation and Artificial intelligence, Nanjing University of Post and Telecommunications, Nanjing 210023 China

[a] chensheng@njupt.edu.cn; [b] daijianbang@yeah.net; [c] 15252769841@163.com

[*] the corresponding author

Keywords: Emotion; Detection; Facial expression; Classroom; Teaching; Feedback

Abstract. Currently, the one-to-many teaching method is still adopted among most of the teaching classroom in modern universities. Limited by the resources of teachers, the number of students in some basic courses could be well into the hundreds. In this teaching situation, it is difficult for teachers to notice learning feedback from all student immediately. This paper proposed a real-time feedback system based on the array camera to obtain facial expressions of all students and judge the teaching effect of the knowledge points. This system can help the teacher handle the learning state of the students dynamically, so that the teaching methods or progress can be adjusted to achieve better teaching effect.

Introduction

Cognitive psychology shows that cognitive processing and emotion processing are integrated at multiple levels in brain [1]. For example, some cortical structures (i.e. orbitofrontal cortex) integrate cognitive and emotion information in learning progress by interacting with neural structures of emotional processing (i.e. amygdala) [2].Drugs can reduce or increase the destructive effect of anxiety on learning[3]. Animal experiment have shown that the learning of the mouse will be affected if the amygdala processing the original neutral sensory signal into an obvious emotional signal with disgusting information [4]. The important role of emotional experience in learning enlightens us that: students can learn and understand more effectively with a classroom atmosphere which can induce positive emotions.

To better induce students' positive emotions, teachers might have to grasp the students' learning states accurately. However, in modern classroom, especially the basic courses, students are often overcrowded, that make it difficult for teachers to monitor students' emotion in learning comprehensively. Nevertheless, with the development of computer vision in recent years, the accuracy of facial expression based on face detection has been continuously improved. Thus, it provides a powerful guarantee for us to build a real-time feedback system of students' classroom learning emotion. Based on the emotional classify of learning emotions, this paper uses camera array to construct image acquisition system; dynamic recognize face information in real time; uses feature extraction and emotion classifier to recognize emotion; summarize identification information to give statistical result and present to teacher in real time.

Emotion and Classify

A general definition of emotion is: emotion is a complex state of feeling that results in physical and psychological and influence mind and body. These changes can also affect thought and behaviors. There are also various types of emotions. Paul Ekman firstly divided emotions into six main categories: Anger, Disgust, Fear, Happy, Sadness, Surprise [5]. He also expanded other emotions such as Shy, Satisfied, Pride, Pleasure, etc. [6] Robert Plutchik [7] summarized eight different emotions and their relationship through wheel of emotion. He Wei [8] established a learner emotion model based on three-dimensional emotion model, and quantified the six explicit expressions. Nowadays, the mainstream expression recognition method is still based on 6+1 classification mode to give comparison result. But for the students' learning statue in classroom situation, to study whether

  DOI: 10.25236/icess.2019.179

these six emotions can help study is still necessary. Base on whether it is conductive to learning, we set up three categories of emotional states, including positive effects on learning, negative effects, and insignificant effects.

Camera Installation

To dynamically identify each student's emotional state, it is required to detect each students' head portrait information in real time. Based on the selected cameras' visual angle, resolution, size of different classroom, student number, the optimal camera installation location can be plan out. If the classroom is too large, an array of cameras can be set. Fig.1 shows the installation layout of two cameras.

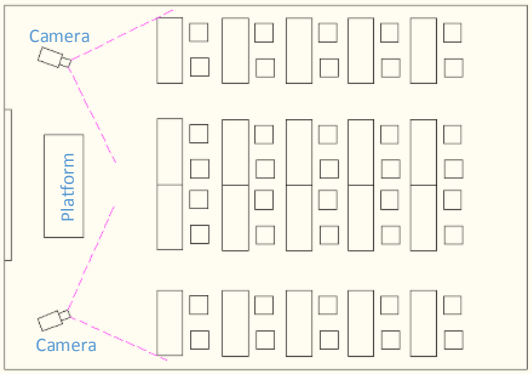

figure 1. Camera Installation Layout

Facial Expression Recognition

Normally, facial expression recognition is consisted of three basic steps. First: face preprocessing; Second: facial emotion feature extraction; Third: emotion classify by classifier. Three steps are shown in Fig 2.

Face detection is a computer technology which provides the location and size information of faces in images. According to the information provided by face rectangle

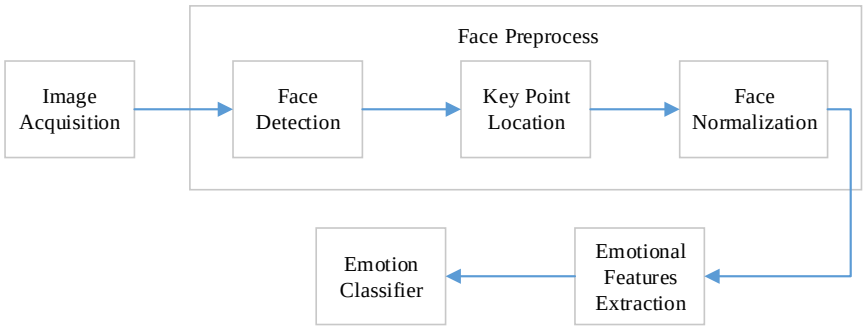

figure 2. Facial Expression Recognition

In face detection, facial feature points can be precisely located. After that face normalization, expression analysis can be applied. Face normalization aligns the face to a predefined template according to the feature points' position. The emotion feature extraction process extracts robust face features from normalized face images. Finally, comparing the extracted features with known emotion features, the emotion can be classified.

Face Preprocessing

According to the installation layout of cameras mentioned above, each camera has no less than one

student in the field of view. In order to accurately identify the emotional state of each student, it is necessary to detect faces in frames. Multiple methods of face detection can be used here: geometric feature-based methods, template matching methods, subspace LDA, neural network-based method, etc.[9] In pre-processing, we also need to normalize, enhance, align the recognized face part according to the background, lighting condition, hand posture and other actual detection situation.

Currently, Viola-Jones and dlib are the two kinds of most widely adopted face detectors, the calculations are also relatively straightforward, but are only effective on front face detection, and still remain some problems in face detection in multiple angles. Some recent works implement face detection by introducing a deformable part model (DPM) [10], which can effectively improve robustness and local accuracy. Fig.3 shows how face are normalized after detected.

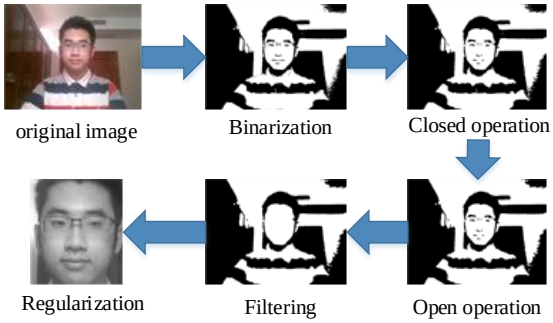

figure 3. Face Normalization

Feature Learning

Feature extraction is the most important step in facial expression system. According to expression features of face, the FACES face classification system proposed by Ekman P [11] gives 32 facial muscle movements and 14 additional motion description. Wang Y and Guan L [12] used 1856 texture feature libraries for pattern matching. In order to acquire these facial features effectively, a Gabor filter and a discrete wavelet transform can be used.

Gabor wavelet transform is a powerful image processing algorithm. It is similar to the perception in the human visual system and can effectively promote image edge detection. Expression classification depends heavily on the shape of facial features such as the mouth, eyes and eyebrows. The kernel function of a Gabor filter is generally defined as [13]:

$$\psi_{u,v}(z)=\frac{\|k_{u,v}\|}{\sigma^2}e^{\left(-\|k_{u,v}\|^2\|z\|^2/2\sigma^2\right)}\left[e^{ik_{u,v}z}-e^{-\sigma^2/2}\right] \tag{1}$$

Where μ and ν defines the direction and proportion of the Gabor filter, and $k_{\mu,\nu}=k_\nu e^{i\varphi_u}$, $\varphi_u=\frac{\pi\mu}{8}$, $k_\nu=\frac{k_{max}}{f^\nu}$ represents the frequency (scale) of the kernel function, f is the interval factor between kernels in the frequency domain, and σ is the standard deviation of the Gaussian kernel function. Normally $\sigma=2\pi, k_{max}=\pi/2, f=\sqrt{2}$.

After obtaining the feature vector of facial expression by Gabor wavelet, a current large facial expression recognition public database fer2013 data set can be used for feature training, which including 35586 face images (containing 28709 training images, 3589 public-test and 3589 private-test images). Each sample in the database has difference in age, facial direction, etc., and has certain practical significance. Features can be extracted layer by layer from simple to complex, by using convolution kernel. The output of muti-layer convolution kernel changes from low-level features, such as the edge of the eyebrows, corner of the eye, to advanced features such as eyes, mouth.

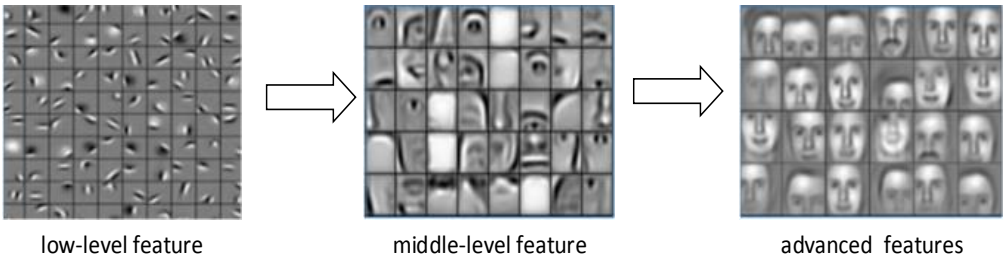

figure 4. Features Exacted by Kernel

MLP Emotion Classifier

Once high-level features are identified by feature extraction, emotions will be classified as seven categories (angry, disgust, fear, happy, neutral, sad, and surprise). A Multi-Layer Perceptron network (MLP) is used in classification. After comparing a series structure parameters of different network, the input layer neurons number is set to be 8400, the 93 neurons are set in hidden layer, and the output layer contains 7 neurons which corresponding to the 7 emotion categories. MLP adopts the sigmoid function and a reverse BP network for training, which can reduce the possibility of falling into local optimum.

In order to further improve teaching measures for various emotional states, emotions are divided into three categories based on the learning statue: positive effects, negative effects, and insignificant effects on learning. With a simple emotion mapping, classification of students' learning state in classroom is realized. sadness, anger, and digest are mapped to the negative state; happy, surprise are mapped to the positive state; neutral and fear are mapped to insignificant/calm state. Meanwhile, face characteristics detection algorithm is used to identify the high-level facial features (i.e. eyes, nose, mouth angles on both sides). A student is considered to be bowed and completely unfocused if less than two features are detected; conscious, looking up, but not looking at blackboard/teacher directly if two or three features are identified; concentrate and look directly at blackboard/teacher if more than four features are detected. Emotion recognition only classifies emotions for faces in which more than four features can be recognized. The classification process is shown in Fig 5.

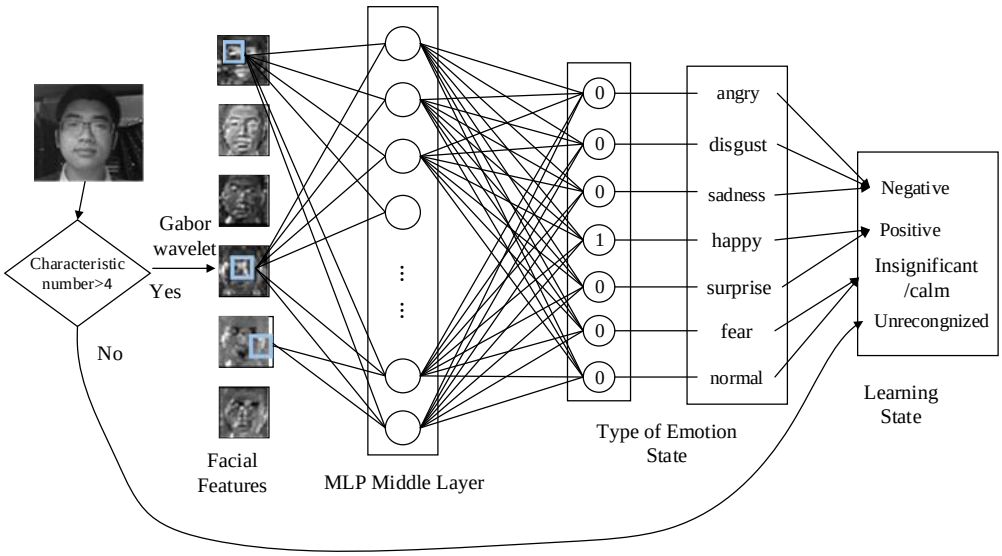

figure 5. Emotion Classify Model Based on MLP

System Construction and Experimental Results

A system test platform is built according to the theory above. Two GS3-U3-51S5M cameras produced by Point Grey are used to collect students' avatar information with a maximum resolution of 2448*2048 at 75fps. The host computer is HP laptop (CPU: intel 17-8550U, memory 8G, Graphics

card: NVIDIA MX150). The self-developed software system is used for real-time monitoring students' emotions. The range is set according to the focal length and resolution of the camera. The number of students whose emotions can be simultaneously identified in one frame is 7 people. The recognition effect is shown in Fig 6.

In a 45-minute course of Modern Test Theory, student's states are tracked and monitored as shown in Fig 7. Fig7(a) shows that students are not in learning state before the class begin, the attention and sensitivity levels are at lowest point; Fig 7(b) shows that students' attention reach the highest peak and their emotion are mostly confused, neutral or insignificant when class reach a key point; Fig 7(c) shows after a difficult knowledge point, the students' attention have a center drop.

(a) Face Detection Result (b) Emotion Detection Result

figure 6. The results of Emotion Detection

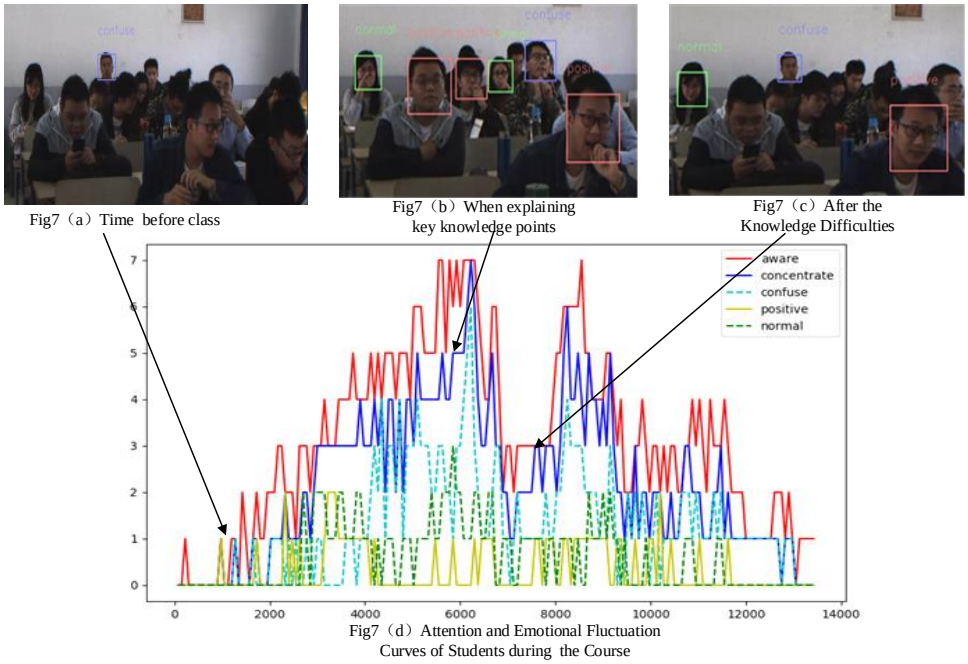

Fig7（a）Time before class

Fig7（b）When explaining key knowledge points

Fig7（c）After the Knowledge Difficulties

Fig7（d）Attention and Emotional Fluctuation Curves of Students during the Course

figure 7. Emotion curve

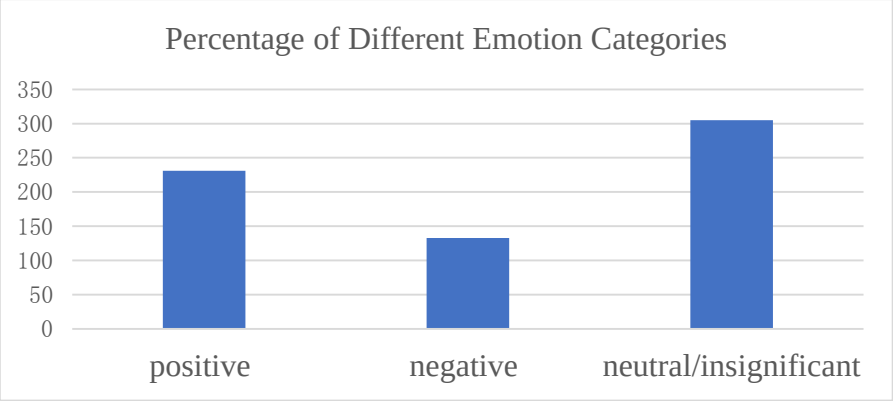

figure 8. percentage of different emotion

It is obvious that in first half of the course, the students' attention gradually increased and emotion are mostly confuse. After high concentrate statue, there was a drop in attention, and the emotion gradually turn to neural. When it comes to the middle and late stage of the class, the students' attention raised again to a significant height with emotion dominated by confusion and neutral. Then attention reduced to a middle level. The percentage of different emotion are shown in fig8. According to changes in students' emotional state, teachers can improve students' attention by using some positive measures to affect students' learning emotions, such as playing animations, asking questions, group discussions.

Summary

This paper introduces a multi-camera-based emotion detection system in classroom environment. The system can detect and record changes in students' facial expression, and report to teacher in real time. By comparing the results of timely and non-timely emotions, it is shown that the result of this system can correctly reflect the real situation. This inspires us that teachers can dynamically adjust the teaching process according to the system. Also it provides reference for making adjustments in teaching plan when students' emotions change, so that the teaching quality can be improved.

Acknowledgements

This research is funded by the Teaching Reform Research Project of Nanjing University of Posts and Telecommunications (JG00518JX58).

References

[1] Kotz, S. A., & Paulmann, S. (2011). Emotion, Language, and the Brain. Language and Linguistics Compass, 5(3), 108–125. doi:10.1111/j.1749-818x.2010.00267.x

[2] Csibra G, Gergely G. Social learning and social cog-nition: The case for pedagogy.[M]. Attention & Perform-ance, 2006, 21: 249－274

[3] Cools R，Roberts A C，Robbins T W. Serotoninergicregulation of emotional and behavioural control proces-ses. [M]. Trends in Cognitive Sciences，2008，12(1) : 31－40

[4] Stutzmann G E，Ledoux J E. GABAergic antagonistsblock the inhibitory effects of serotonin in the lateralamygdala: A mechanism for modulation of sensoryinputs related to fear conditioning．[J]. Journal of Neuro-science，1999，19(11) : RC8

[5] Ekman P. Universals and cultural differences in facial expressions of emotion.[J]. Journal of Personality & Social Psychology, 1987, 53(4):712.

[6] Ekman P. Basic Emotions.[M]. Handbook of Cognition and Emotion. John Wiley & Sons, Ltd, 2005:45—60.

[7] Thamm R A. The classification of emotions.[M]. Handbook of the sociology of emotions.

Springer, Boston, MA, 2006: 11—37.
[8] Hugo Lövheim, H. A new three-dimensional model for emotions and monoamine neurotransmitters. Medical Hypotheses, 78(2), 341–348. doi:10.1016/j.mehy. 2011.11.016
[9] Phillips, P. J., Flynn, P. J., Scruggs, T., Bowyer, K. W., Chang, J., Hoffman, K., … Worek, W. (n.d.). Overview of the Face Recognition Grand Challenge. 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). 2005. 268
[10] Orozco J , Martínez B, Pantic M. Empirical analysis of cascade deformable models for multi-view face detection.[J]. Image & Vision Computing, 2015, 42(C): 47—61.
[11] Ekman P, & Rosenberg E L. What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS).[M]. Oxford University Press, USA, 1997:425—434.
[12] Wang Y, Guan L. Recognizing Human Emotional State From Audiovisual Signals.[J]. IEEE Transactions on Multimedia, 2008, 10(5): 936—946.
[13] Chelali F Z, Djeradi A. Face Recognition Using MLP and RBF Neural Network with Gabor and Discrete Wavelet Transform Characterization: A Comparative Study.[J]. Mathematical Problems in Engineering, 2015: 1—16.