

Emotion Analysis Based on Facial Expression Recognition in Virtual Learning Environment

Yongna Liu¹, Lirong Wang², Wanping Li^{1*}

¹ College of Applied Science and Technology, Hainan University, Hainan, China, 571737.

² Department of Education, Xinzhou Teachers University, Shanxi, China, 034000.

* Corresponding author. Email: vicsin@sohu.com

Manuscript submitted March 2, 2016; accepted July 30, 2016.

doi: 10.17706/ijcce.2017.6.1.49-56

Abstract: The psychologic research find that learners' cognitive activitie had been affect by emotion in their learning process, specifically, positive emotion can promote cognitive activities, whereas negative emotion can hinder them. The affective state can be seen as a very strong predictor of thinking level in learning. Therefore, establishment of a intelligent environment and atmosphere which can recognize and adjust the emotion of students plays an important role in positive learning. Application of facial expression recognition is performed on Magic Learning which is a 3D Virtual Learning Environment (3DVLE) developed by our laboratory and experimental results suggest that our method is feasible and effective.

Key words: Emotion recognition, smart learning, virtual learning environment (VLE).

1. Introduction

Psychologist Mehrabian said that when people put forward an idea, there are three elements link the hearer with the speaker: words constitute 7%, tone of voice constitute for 38%, and facial expressions constitute the biggest and the last part, 55%. From this, it can be seen that as the main way of human express their emotion, expression contains lots of information about the inner emotional changes. According to facial expression, people can infer the inner subtle emotional state accurately.

In VLE, we not only pay attention to virtual teachers' emotional expression, but also need to focus on the feedback—learners' emotional state in learning process. We can collect the learners' inner emotional state by obtain their facial expression, for example: collect their faces images by using the camera of mobile devices. Compare with galvanic skin response and brain waves, facial expression is much easier to obtain and analysis because of this method do not require complicate and precision instruments, it is more viable in daily work. In fact, through facial expression to analysis emotion is underway in learning processing.

2. Facial Expression Recognition in VLE

2.1. The Framework of VLE-FER

The framework consists of five layers: sensing layer, transmission layer, data layer, analysis layer, and application layer. The mobile terminals (such as webcam, handset, iPad, and laptop etc.) in sensing layer are used for capturing the video of learners in learning process. These devices allow us to collect and capture the data of learners wherever they are. Transmission layer can provide a powerful internet environment for us to ensure the data transmitting quickly and securely. The video data and facial expression images are

stored in data layer which it based on performance and reliability requirements. The main function of analysis layer is facial expression analysis, which is the most important part in VLE-FER. Specifically, when considering facial expression recognition based on image it includes three procedures: face feature extraction, facial expression feature decomposition, and facial expression classification. Then, we can obtain the affective state from classification information of facial expression image. Using affective state we can evaluate and analysis the learning behavior from the emotional dimension, such as the concentration degree, patience degree, understanding degree and fatigue degree. This would be able to solve the problem that was not possible to analysis from cognitive dimension.

2.2. Affective State and the Corresponding Facial Expression in Learning

At present, six basic emotions identified by Ekman and Friesen have been widely accepted, they are happy, sad, surprise, angry, disgust, and fear. However, many researches have shown that using these emotions to research the emotional states of learners in their learning processing is unreasonable because they do not play an important role in learning. Kort's research has found that a set of affective states typically play a significant role in learning, particularly in the case of complex learning of college students, they are boredom, confusion, frustration [1]. It is worth noting that the affective states mentioned above with different frequencies in learning. Craig's research suggests that delight and surprise are the most frequently occur in learning [2]. Therefore, the affective states of delight, confusion, boredom, surprise, and fatigue in learning will be primary concerned in this paper.

3. Feature Extraction

3.1. Prior Work on Feature Extraction

In recent years, how to use facial expression recognition technology analysis learners' affective state during on-line study has become to one of the hottest research topic. One of the most successful techniques to facial feature extraction is the subspace method, such as Principal Component Analysis (PCA) [3]. However, according to resent research, the face images possible reside on a nonlinear sub-manifold. Unlike many existing techniques such as PCA, manifold learning describes the manifold structure by a nearest-neighbor graph which preserves the local structure of the image space. The algorithms of manifold learning are now available for feature extraction, such as LPP which is the linearization method of LE. Currently, LPP is widely used in the field of face recognition.

However, the feature extracted by the method above is the combination of individual human facial feature and expression feature. In an optimal situation, the related individual facial feature can be separated during the process of facial expression recognition. Higher-Order Singular Value Decomposition (HOSVD) [4] is a general n-factor analysis method. HOSVD has been applied to human face recognition and facial expression recognition. We use a third-order tensor $D \in R^{i \times j \times k}$ to represent the facial expression configuration. The goal is to decompose tensor using HOSVD to extract separate subspaces along person mode and expression mode.

3.2. Facial Feature Extraction

We aim to find an objective function that can preserve the manifold structure of the face. The method of extracting facial feature absorbs the advantages from LPP.

LPP [5] seeks a low-dimensional representation to retain local structure of the original data. Given a set of data points $\{x_i\}_{i=1}^n$, where $x_i \in R^m$ is an m-dimensional column vector. The objective function is defined as:

$$\min \sum_{ij} (y_i - y_j) W_{ij} \quad (1)$$

where y_i is the one-dimensional representation of x_i , and the matrix $W = \{w_{ij}\}_{n \times n}$ is a similarity matrix. W is defined as:

$$w_{ij} = \begin{cases} \exp\left(-\|x_i - x_j\|^2 / \delta\right), & \text{if } x_i / x_j \text{ is among } KNN \\ & \text{of } x_j / x_i \\ 0, & \text{otherwise} \end{cases}$$

4. Feature Decomposition

In this section, facial feature we obtained using LPP will be divided. Specifically, we use Higher-Order Singular Value Decomposition (HOSVD) [4] to decompose the facial feature and the expression feature into face subspace and the expression subspace respectively. In this way, the interference of facial features can be eliminated completely when we recognize the facial expression.

4.1. Basic Knowledge of Tensor

4.1.1. Higher-order singular value decomposition

The visualization of the higher-order singular value decomposition for third-order tensor in Fig. 1.

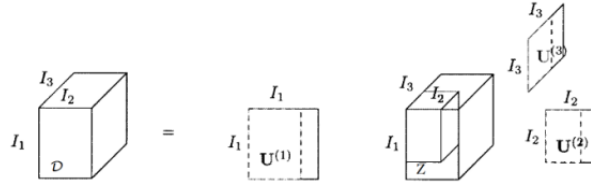


Fig. 1. HOSVD of a third-order tensor [4].

The higher-order singular value decomposition of a tensor $D \in R^{I_1 \times \dots \times I_N}$ with N-order is:

$$D = Z \times_1 U^{(1)} \times_2 U^{(2)} \dots \times_N U^{(N)} \quad (6)$$

where Z is the core tensor, $U^{(n)} \in R^{I_n \times I_n}$, $n(n=1, \dots, N)$ are the orthogonal matrix. The algorithmic procedure of HOSVD is stated in Table 1.

Table 1. HOSVD Algorithm

Algorithm 1	HOSVD
Input: tensor $D \in R^{I_1 \times \dots \times I_N}$; the number of order N .	
Output: core tensor $Z \in R^{I_1 \times \dots \times I_N}$; orthogonal matrix $U^{(n)} \in R^{I_n \times R_n}$, $(n=1, \dots, N)$.	
for $n = 1, \dots, N$ do	
Step 1: unfolding the tensor D $M^{(n)} = \text{unfold}(D), (n=1, \dots, N)$;	
Step 2: singular value decomposition of $M^{(n)} = U_n \sum V_n^T$;	
Step 3: finding the rank of $M^{(n)}$ $R_n = \text{rank}(M^{(n)})$;	
Step 4: finding first R_n left eigenvectors $U^{(n)}$ in U_n , let U_n equal to $U^{(n)}$, $U^{(n)} = U^{(n)}$;	
end	
Step5: computing core tensor Z $Z = D \times_1 U^{(1)T} \times_2 U^{(2)T} \dots \times_N U^{(N)T}$.	

4.1.2. Facial Expression Decomposition

A third-order tensor will be used to represent the image which is composed of different faces and different facial expressions. By applying HOSVD (as Table 1) the facial feature and facial expression feature can be decomposed into different subspaces. The specific steps are described in the following :

A. 3D Modeling

Given a set of facial expression images of different persons. We use a third-order tensor $D \in R^{i \times j \times k}$ to represent the facial expression configuration, where i is the number of expression for each person, j is the number of person and k is the dimension of the facial feature vector which is use to describe the distribution of gray-level appearance variations. Such third-order tensor $D \in R^{i \times j \times k}$ is shown in Fig. 2.

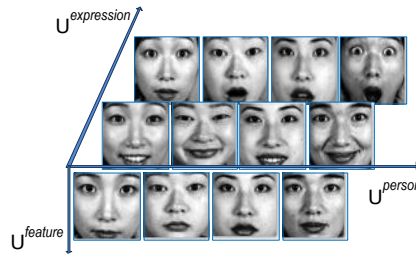


Fig. 2. Third-order tensor.

B. Face Tensor Decomposition

We can decompose face tensor $D \in R^{i \times j \times k}$ into different subspaces using Algorithm 1, $D = Z \times_e U^{expression} \times_p U^{person} \times_f U^{feature}$, where Z is the core tensor which represents the interactions of the expression, person and feature. Matrix $U^{expression}$, U^{person} and $U^{feature}$ represents the expression subspace, person subspace and feature subspace respectively, and these matrices are all orthogonal. Each row vector in each subspace matrix represents a specific vector in the mode. For example, the row vector u_n^e in expression subspace $U^{expression} = [u_1^e, u_2^e, \dots, u_n^e, \dots, u_i^e]^T$ represents the characteristics of the n -th expression. The row vector u_n^p in person subspace $U^{person} = [u_1^p, u_2^p, \dots, u_n^p, \dots, u_j^p]^T$ represents the characteristics of the n -th person. Each column vector in each subspace matrix represents the contributions of other modes. Is worth noting that each column vector in feature subspace $U^{feature}$ represents Eigen Face, while $T = Z \times_f U^{feature}$ represents the tensor-face. By using tensor face we can clearly describe the effects of various factors on image. Such as, an image d_{p_n, e_n} can be expressed as $T \times_p u_n^p \times_e u_n^e$. Therefore, the second kinds of expression of the first person's image can be expressed as $T \times_p u_1^p \times_e u_2^e$.

C. Facial expression recognition

a) Model Training

As mentioned above, the row vector u_n^e in expression subspace $U^{expression}$ can represents the characteristics of the n -th expression. So the coefficients to distinguish different expressions can be obtained through training model using different expression's images, as described in the following procedure:

The base tensor is defined as:

$$B_p = T \times_p U^{person} \quad (7)$$

where, T is the tensor-face. Then, the base tensor of specific individual can be obtained:

$$B_n^p = T u_n^p \quad (8)$$

where, u_n^p is the characteristics of the n -th individual. Therefore, one training sample d_{p_n, e_n} can be expressed as follows:

$$d_{p_n, e_n} = (B_{p_expression})^T c_n^e \quad (9)$$

where, $B_{p_expression}$ is the matrix which can obtained by unfolding tensor B_n^p in expression mode. The coefficient c_n^e can describe characteristics between different expressions, it is called the expression coefficient. According to (9) available:

$$c_n^e = (B_{p_expression})^{-T} d_{p_n, e_n} \quad (10)$$

where, the expression coefficient c_n^e is the vector we aim to find.

b) Facial expression recognition

One test sample can be seen as a tensor $t \in R^{1 \times 1 \times I_3}$. Then, using (11) we can obtain the coefficient of expression c^e . Finally, we calculate the value of $c^e - c_n^e$ and find the minimum value of the corresponding training sample c_n^e , where n is the classification of the expression which we aim to find. So, the test sample's expression coefficient has been obtained.

$$c^e = (B_{p_expression})^{-T} t_{p,e} \quad (11)$$

5. Facial Recognition in Magic Learning

In this section, several experiments are carried out in sub-module of Magic Learning -affective interaction subsystem. Magic Learning is a three dimensional VLE (3DVLE) which is developed by our laboratory [6], [7]. The affective interaction subsystem can be seen in Fig. 3.

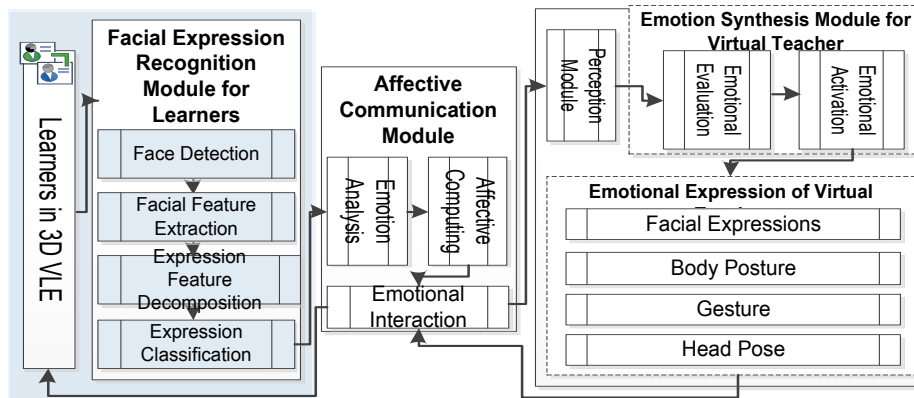


Fig. 3. Emotional interaction framework between learner and virtual teacher in 3DVLE.

In our experiment, ten learners' facial expression videos were captured at random during their learning process. Mobile-phone, laptop, and iPad were used to collect the videos of learners. Fig. 4 shows partial preprocessed facial expression images of learner. These images include four kinds of expressions: delight, concentration, confusion and fatigue.



Fig. 4. Partial processed facial expression images of learner.

Just as illustrated in Fig. 3, the facial expressional recognition result will as the input variable of the affective interaction module. Then after emotion analysis and affective computing, the corresponding emotion expression of virtual teacher will be activated [8]-[10], see Fig. 5. When learner shows delight, virtual teacher will smile, nod her head, or shows other emotional expression of praise; When learner shows fatigue, virtual teacher will intervene, like shake her head; When learner shows confused, virtual teacher will encourage the learner, such as pat him on the shoulder. The emotional expression of virtual teacher is shown in Fig. 6.

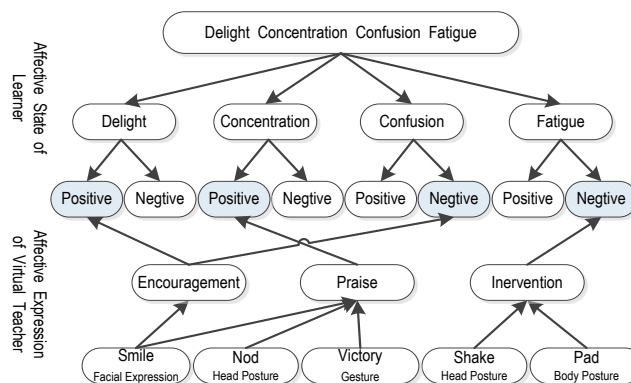


Fig. 5. Affective interaction between virtual teacher and learner.



(a) nodding and tossing



(b) smiling, tapping someone on the shoulder, and the victory sign

Fig. 6. Emotional expression of virtual teacher.

These experiments activate virtual teacher's emotional expression by facial expression alteration of learners, thus realizing the emotional interaction in 3DVLE. Specifically, learners will be encouraged by

virtual teacher when they meet with difficulties in learning and will be praised by virtual teacher when they excellent in learning. In this way, the negative emotion of learners can be adjusted rapidly and the positive emotion of learners can be maintained in their learning process.

6. Conclusion and Future Work

The study about VLE that can apply learners' facial expression recognition is an important direction for future smart learning development. Facial expression feature extraction is an important part for facial expression recognition, because it determines the effect of the expression recognition directly. The feature extraction method presented in this paper is significant in improving the human-machine affective interactive level. However, the practical complicate application environment such as posture, lighting and blocking, will undesirable impact on the expression recognition rate. What' more, using facial expression recognition techniques for affective analysis requires facial expression database which contains the facial expressions of learners in learning environment rather than the only six basic facial expressions studied widely in the fields of expression recognition. Therefore, these problems need to continue in-depth studies.

Acknowledgment

This work was funded by Natural Science Foundation of Hainan Province of China (20166212), Natural Science Foundation of Hainan Provincial Department of Education (Hnky2016-5) and Scientific Research Fund of Hainan University (kyqd1619).

References

- [1] Kort, B., Reilly, R., & Picard, R. (2001). An affective model of interplay between emotions and learning: Reengineering educational pedagogy — building a learning companion. *Proceedings of IEEE International Conference on Advanced Learning Technology: Issues, Achievements and Challenges* (pp. 43-48). Madison, Wisconsin: IEEE Computer Society.
- [2] Craig, S., Graesser, A., Sullins, J., & Gholson, B. (2004). Affect and learning: An exploratory look into the role of affect in learning. *Journal of Educational Media*, 29, 241-250.
- [3] Kshirsagar, V. P., Baviskar, M. R., & Gaikwad, M. E. (2001). Face recognition using eigenfaces. *Proceedings of IEEE Conf. Computer Research and Development* (pp. 302-306).
- [4] Lathauwer, L. D., Moor, B. D., & Vandewalle, J. (2000). A multilinear singular value decomposition. *Society for Industrial and Applied Mathematics Journal of Matrix Analysis and Applications*, 21(4), 1253-1278.
- [5] He, X., Yan, S., Hu, Y., Niyogi, P., & Zhang, H. (2005). Face recognition using laplacianfaces. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 27(3), 328-340.
- [6] Zhao, H., & Sun, B. (2009). Emotional learning environment construct based on virtual reality technology under the network (in Chinese). *China Educational Technology*, 4, 101-104.
- [7] Zhao, H., & Sun, B. (2012). Research on emotion synthesis model for virtual teacher (in Chinese). *China Educational Technology*, 1, 121-126.
- [8] Zhao, H., & Sun, B. (2010). Research on 3D agent-based virtual teacher with emotional supports (in Chinese). *China Educational Technology*, 11, 117-123.
- [9] Zhao, H., Hu, X., Xie, B., & Tian, Y. (2011). Emotion expression base on body language for 3D virtual teacher (in Chinese). *Computer Engineering*, 37(23), 159-164.
- [10] Sun, B., Xie, B., Zhao, H., Yao, L., & Zhu, X. (2011). Research and improvement on character animation system of 3D virtual learning platform. *Proceedings of African Conference on Software Engineering and Applied Computing* (pp. 65-71). Cape Town, South Africa.



Yongna Liu received the PhD degree in educational technology from Beijing Normal University, Beijing, China in 2015. She is now a lecturer at Hainan University. Her current research focuses on learning emotion recognition. The aim of her work is to better understand the relationship between inner emotion and facial expression in learning and to find feasible and effective emotion recognition techniques. Her research interests include affective computing, image processing and facial expression recognition.



Lirong Wang is an assistant of the Department of Education, Xinzhou Teachers University. Her main research interest is distance education, the domain aimed at application of virtual reality technology in education.



Wanping Li is an associate professor of the College of Applied Science and Technology at Hainan University. Her main research interests include pattern recognition and applications of new technologies about Internet of things.