Emotion Evaluation Through Body Movements Based on Silhouette Extraction

Hong Yuan[1,2], Bo Wang[1(✉)], Li Wang[1], and Muxun Xu[2]

[1] National Laboratory of Human Factors Engineering,
China Astronaut Research and Training Center, Beijing 100094, China
yuanhong0624@126.com, wowbob@139.com, wanglikunyu@126.com
[2] Department of Industrial Design, Xi'an Jiaotong University, Xi'an 710049, China
xumuxun@xjtu.edu.cn

Abstract. In the modern age of technology development, the ability of computer to recognize and express emotions are expected in Human-Computer Interaction. This paper proposed a method to evaluate human emotions by extracting human explicit behaviors during interaction, which prevented interference from outside. To achieve the aim, OpenCV was used to extract silhouette features of physical behaviors and PAD scale was used to evaluate the emotions on time. And then, Emotion-Behavior Library was established based on both silhouette features and the results of PAD scale. In order to obtain positive, negative and negative emotions, Chinese folk music were used as emotional material in the experiment. Subjects were asked to express their emotions by body movements. The results verified the hypothesis: (1) behavioral features can effectively represent emotions; (2) the method for extracting the behavioral characteristics in this study is effective. Besides, we found that the front camera has a higher accuracy, for example accuracy of positive and negative was 72.5%, while the data of right camera hasn't been classified effectively.

Keywords: Emotion evaluation · Physical behavior · PAD scale · Affective computing · Emotional classification

1 Introduction

The communication between people is not only rational and logical, but also natural and emotional. In the modern age of technology development, the ability of computer to recognize and express emotions are expected in Human-Computer Interaction (HCI). As an important branch of HCI, Affective Computing is related not only to psychology of emotion, but also to computer technology and statistical analysis [1]. Human emotions are implied in explicit features which play an important role in HCI, such as verbal behaviors, facial expressions and physical behaviors [2]. With the maturity of computer vision and machine learning, the analyses of individual behaviors and emotions become more and more popular.

According to previous research, the emotional recognition of behavior was usually associated with static gestures or inflexible limited behaviors [3, 4]. This paper proposed

a method to evaluate human emotions while the interaction is in progress, which prevented interference. We focus on real-time emotional evaluation through human explicit behavior. To achieve the aim, OpenCV was used to extract silhouette features of human behaviors and (PAD) emotional scale was used to evaluate the emotions on time. We formed a model between behaviors and emotions.

2   Related Work

The research of affective computing in HCI includes facial expression, speech emotion, body movement, text emotion and so on [2, 5, 6]. And previous researchers have proposed plenty of methods to evaluate human emotions. Ekman and Friesen studied the relationship between emotional strength and posture in 1967 [7]. Camurri et al. described the features of body movements through physical parameters, such as the location and speed of the body [8], while Castellano found that emotional recognition of body movements was more effective than verbal language and facial expressions [9]. Compared with the method of subjective evaluation, extracting physical behavior was non-intrusive and less interference, which made the abstraction of data more convenient and objective [8, 9].

With the development of information technology, researchers tried to obtain human behaviors with wearable sensors and cameras, and to analyze human behaviors through computer technology [10]. The way of obtaining human behaviors with wearable sensors was more direct and accurate, but more interferential. However, the method of cameras has the advantages of non-invasive, convenient, but confined by the background and target movements [11].

Emotional recognition based on behaviors was mainly to establish an emotion-behavior library by discriminating and analyzing the characteristics of various movements, which were extracted from the movement features of the body under various emotional states [3, 4, 12]. Human movement features, such as the duration, frequency and other properties, were the basis of emotion recognition. Laban and Ullman proposed the Body Action Coding System (BACS) containing information: the part of the body (such as the left hand), direction, speed and shape (such as hand clenched into a fist), distance (curve/line), strength (weak/strong), time (continuous/fast) and fluency [13]. However, most of postures or movements do not have obvious emotional characteristics, and can't be used fully identify the resolution.

Under the condition of laboratory, emotional induction was an effective way to obtain effective emotional reaction. In the past few decades, researchers put forward lots of methods of to induce emotions including the methods of recall, imagination, image, film and music [14]. The methods of recall and imagination had instable result, while the duration was short induced by picture. The method of film had complex factors, while the music was immersive, coincident and duration long [14, 15]. In previous study, the music was mostly Western classical music, which has different effects on the Chinese people's emotion because of cultural differences [15]. In this paper, we chose "Chinese folk music emotional library" as the emotional induction material, which was established

in our preliminary study. The music library can effectively induce positive, neutral and negative emotions, which are consistent in different subjects.

The PAD scale was used to evaluate the emotion in both this paper and the establishment of "Chinese folk music emotional library". The theory of discrete emotion believed that human emotion was discrete and measurable [16]. On the other hand, the 6 basic emotions, involving anger, disgust, fear, happiness, sadness and surprise [17], proposed by Ekman was applied widely. The theory of continuous emotion held that human emotion was continuous, complex and distributed in a certain range. The PAD model proposed by Mehrabian and Russell consisted of three dimensions: pleasure-displeasure, arousal-nonarousal and dominance-submissiveness [18]. Continuous emotion model was convenient for feature modeling, which can cover the typical types of discrete emotions [16].

3 Methods

3.1 Features Extraction

Silhouette features, involving the silhouette area of the target: *Area* (*Silhouette*[t]), the smallest external polygon area of the target: *Area* (*MinPolygon*[t]), and the silhouette centroid coordinate(x[t] and y[t]), were extracted through OpenCV function. And then, the extracted data was cleaned by removing the non-target data and filling the missing frame.

The features extracted from the previous step were calculated and sorted into 3 categories of parameters. There were two state parameters: relative state of silhouette area (*CI*[t]), and the barycentric coordinates of silhouette (x[t] and y[t]). And there were four change rates of parameters: the change rate of silhouette area (*RoSC*[t]), the change rate of barycentric coordinates (v[t]), the change rate of silhouette area (A[t]), and the change rate of barycentric coordinates (a[t]). The time series data (the state parameters, the change rate parameters and the first derivative of change rate) were normalized by extracting the parameters of each time series data.

$$\text{RoSC}[t, n] = \frac{\left(\sum_{i=t}^{t+n} Silhouette[i] \right) - \left(\sum_{i=t-n}^{t} Silhouette[i] \right)}{\sum_{i=t-n}^{t} Silhouette[i]} \tag{1}$$

$$CI[t] = \frac{Area(Silhouette[t])}{Area(MinPolygon[t])} \tag{2}$$

3.2 PAD Scale

PAD emotional scale was designed based on PAD emotional state model consisting of the dimensions of Pleasure, Arousal and Dominance. The PAD scale used in the study was simplified Chinese version revised by Researchers at the Institute of psychology Chinese Academy of Sciences based on the scale proposed by Mehrabian. Each dimension of the scale is measured by 4 items, which can be used to evaluate and classify

human emotions effectively. In this study, after listening to each piece of music, subjects were asked to evaluate the scale according to their actual emotional state. More consistent the state to their emotional state more the score is close to 4, or else to −4 (see Table 1).

Table 1. Emotional scale (It was in Chinese in the experiment)

	Score	
Annoyed	−4 −3 −2 −1 0 1 2 3 4	Pleased
Sleepy	−4 −3 −2 −1 0 1 2 3 4	Awake
Powerless	−4 −3 −2 −1 0 1 2 3 4	Powerful
Scornful	−4 −3 −2 −1 0 1 2 3 4	Friendly
Calm	−4 −3 −2 −1 0 1 2 3 4	Stimulated
Submissive	−4 −3 −2 −1 0 1 2 3 4	Dominant
Sad	−4 −3 −2 −1 0 1 2 3 4	Happy
Relaxed	−4 −3 −2 −1 0 1 2 3 4	Active
Humble	−4 −3 −2 −1 0 1 2 3 4	Important
Enraged	−4 −3 −2 −1 0 1 2 3 4	Excited
Bored	−4 −3 −2 −1 0 1 2 3 4	Surprised
Influenced	−4 −3 −2 −1 0 1 2 3 4	Influential

3.3   Materials

Set-Up. The experiment was held in Beijing. The size of the experimental site was 4×4 m (Fig. 1). Two cameras were settled in front of the site and the right. Ten subjects participated in the experiment.

a. front camera   b. right cammera

Fig. 1. Experimental stage and a subject

Emotional induction. In this study, Chinese folk music was used as the material of emotional induction, which has good immersion, persistence, and validity. A total of 21 pieces of music were chosen from "Chinese folk music emotional library", including positive, neutral and negative music each 5 pieces and from which 2 pieces were picked from each category (for test-retest reliability of the data) (see Table 2).

Table 2. Number of induction material

	Positive	Neutral	Negative
Original music	5	5	5
Repeated music	2	2	2
SUM	21		

3.4 Experimental Procedure

Firstly, emotions involving positive, medium and negative were induced by 21 periods of Chinese folk music, while 10 subjects were asked to express their emotions through body movements while listening music. Secondly, the movements of subjects were recorded by 2 cameras, the position of which were front and right of the stage. After each piece of music, the emotions of subjects were evaluated through PAD scale for a few seconds. Then, the silhouette parameters of subjects in the video were extracted through our methods. Finally, the training models were established with the parameters and PAD scales, which was used to evaluate the unknown emotions implied in body movements (see Fig. 2).
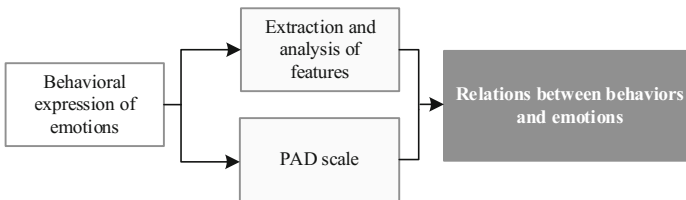
Fig. 2. Experimental procedure

4 Result Verification

4.1 Data Processing

Emotional Scale

Reliability. Software of SPSS 22 was used for data statistical analysis. Internal Consistency Coefficient of PAD emotional scale: $\alpha = 0.888$; for test-retest reliability, independent sample T of repeated induction material: $P = 0.104$. There was no significant difference between repeated induction, which indicated the consistence of perception and evaluation of subjects.

Classification. The emotions of the experimental music were divided into 2 major categories or 3 specific categories through clustering analysis (see Table 3) of PAD scale. The result of clustering analysis was consistent with the hypothesis: the first 7 pieces of music corresponded to negative emotion, the middle 7 pieces of music corresponded to neutral emotion, and the last 7 corresponded to positive emotion.

Table 3. Classification accuracy of each category

Classification	Details	Pieces of music
Type 1	Positive	15,16,17,18,19,20,21
	Negative	1,2,3,4,5,6,7,8,9,10,11,12,13,14
Type 2	Positive	15,16,17,18,19,20,21
	Neutral	8,9,10,11,12,13,14
	Negative	1,2,3,4,5,6,7

Silhouette feature data. The OpenCV was used to extract the features of the 10 subjects in each of the 21 videos, which described the body characteristics of the targets. Then the data was cleared and the missing values were processed. Finally, the normalized parameters of time series data were extracted, including the mean, median, variance, maximum, minimum and maximum of the sequence.

4.2   Emotion Recognition

Induction material was divided into 2 big categories based on the PAD scale, which could be described as negative and positive emotions; and into 3 categories, which could be described as positive, neutral, and negative emotions. The accuracy of prediction equals the correct number of recognition divided by the total number of samples and multiplied by 100%.

Table 4 lists the corresponding classification accuracy of each camera: the front camera has a high accuracy, for example accuracy of positive and negative was 72.5%; the data of right camera hasn't been classified effectively.

Table 4. Classification accuracy of each category

Camera	Classification	Accuracy
Front	Positive, neutral and negative	50.00%
Front	Positive and negative	72.50%
Right	Positive, neutral and negative	33.33%
Right	Positive and negative	52.50%

5   Conclusion and Discussion

The results verified the hypothesis: (1) behavioral characteristics can effectively represent emotion; (2) the method for extracting the behavioral characteristics in this study is effective in this study.

The internal consistency reliability and the test-retest reliability of the PAD scale was valid. Therefore, the PAD scale was a reliable and effective method for evaluating emotions. 3 categories of the classification results of PAD scale, involving positive, neutral and negative emotions, were completely consistent with the classification of the induction materials, which verified the effectiveness of the induction material.

We found that the expression of negative emotion was close to neutral. The results of classification and recognition verified the conjecture in the cluster analysis: the difference of expression between negative and neutral emotion was small, while the difference between negative and positive emotions was large.
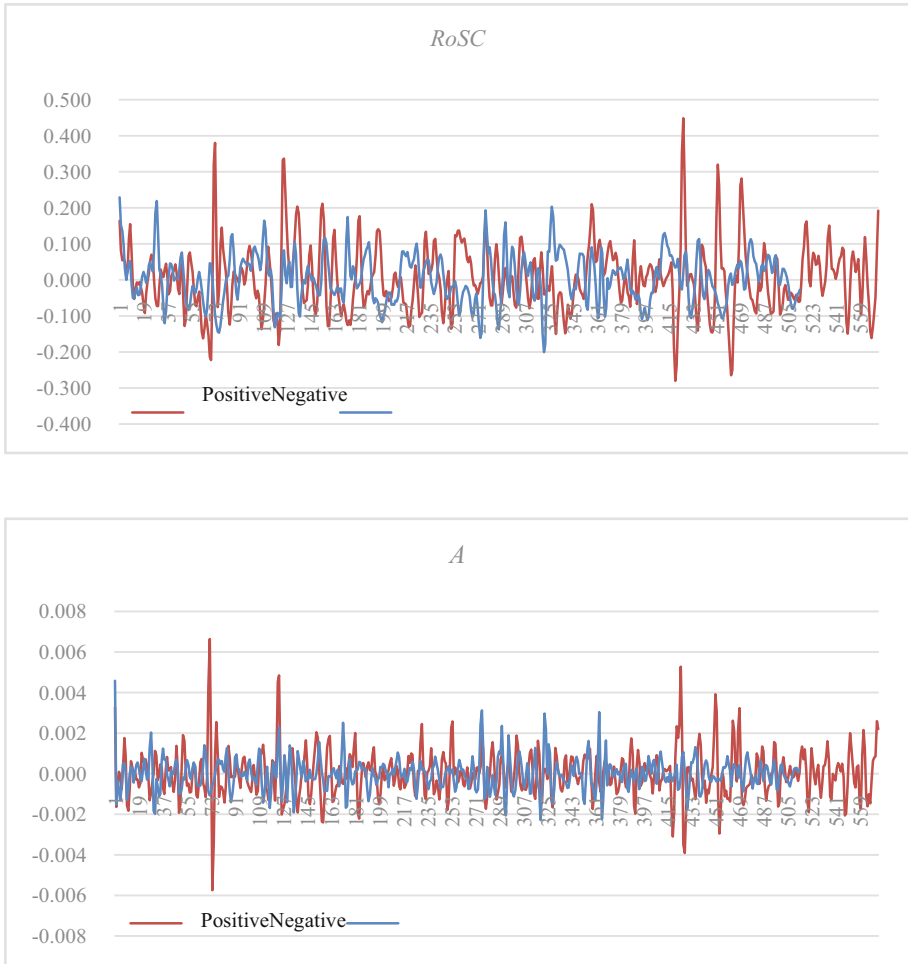

Fig. 3. *RoSC* and *A* of positive and negative emotions (The figure above is the rate of the target's movement, while the below is a derivative of the rate. Blue line represents negative emotion, while red indicates positive emotion) (Color figure online)

Table 5 presents the correlation test of correlation ($0.3 < $ pearson $ < 0.5$) between the parameters and categories. These parameters contributed greatly to the classification results, such as the variance of *A*, the variance of *RoSC* and so on.

Table 5. Correlation between parameters and categories

Parameters	Pearson	Sig.
Variance of A	.404	.000
Variance of $RoSC$	.384	.000
Mean of v	.347	.000
Median of v	.332	.000
Mean of $RoSC$	.317	.000
Number of peaks of v	.304	.000

Through analyzing the characteristics of different emotions, we found that the positive emotion was more powerful than the negative emotion. Speed and strength in BACS corresponded to $RoSC$ and A in Fig. 3. The amplitude of red line is larger than the blue, which means the positive emotion contains more energy.

References

1. Picard, R.W.: Affective computing for HCI. Human-computer interaction: ergonomics and user interfaces. In: Proceedings of HCI International, pp. 829–833. DBLP (1999)
2. Derya, O.: Towards Social Virtual Listeners: Computational Models of Human Nonverbal Behaviors. University of Southern California (2014)
3. De Silva, P.R., Bianchi-Berthouze, N.: Modeling human affective postures: an information theoretic characterization of posture features. J. Comput. Anim. Virtual Worlds **15**(3–4), 269–276 (2004)
4. Kapur, A., Virji-Babul, N., Tzanetakis, G., et al.: Gesture-based affective computing on motion capture data. In: Proceedings of the 1st International Conference on Affective Computing and Intelligent Interaction, pp. 1–7 (2005)
5. Chunling, L., Beiji, Z., Lei, W.: Double-mode estimation of emotion intensity based on facial and action's expression. J. Syst. Simul. **21**(16), 5047–5052 (2009)
6. Ding, Y., Fu, X.: The research of affective recognition on the human motion tracking. Comput. Knowl. Technol. **7**(11), 2649–2651 (2011)
7. Ekman, P., Friesen, W.V.: Head and body cues in the judgment of emotion: a reformulation. Percept. Mot. Skills **24**(3), 711–724 (1967)
8. Camurri, A., Lagerlöf, I., Volpe, G.: Recognizing emotion from dance movement: comparison of spectator recognition and automated techniques. Int. J. Hum. Comput. Stud. **59**(1–2), 213–225 (2003)
9. Castellano, G., Kessous, L., Caridakis, G.: Emotion recognition through multiple modalities: face, body gesture, speech. In: Peter, C., Beale, R. (eds.) Affect and Emotion in Human-Computer Interaction. LNCS, vol. 4868, pp. 92–103. Springer, Heidelberg (2008). doi: 10.1007/978-3-540-85099-1_8
10. Zhang, Y., Lin, X.: Affect is computable-a survey on affective computing. Comput. Sci. **35**(5), 5–8 (2008)
11. Xiaolan, F.: Psychology of Emotion. East China Normal University Press, ShangHai (2016)

12. Shaarani, A.S., Romano, D.M.: Perception of emotions from static postures. In: Paiva, A.C.R., Prada, R., Picard, R.W. (eds.) ACII 2007. LNCS, vol. 4738, pp. 761–762. Springer, Heidelberg (2007). doi:10.1007/978-3-540-74889-2_87
13. Laban, R., Ullmann, L.: The Mastery of Movement, 4th revision edn. Princeton Book Company Publishers, Princeton (1988)
14. Jiang, J., et al.: Mood induction procedures and the recent advancement. J. Southwest China Normal Univ. (Nat. Sci. Ed.) **36**(1), 209–214 (2011)
15. Dongdong, L., Zhenbo, C., Ruina, D., et al.: Preliminary establishment and assessment of affective music system. Chin. Ment. Health J. **26**(7), 552–556 (2012)
16. Lade, P.: Probabilistic Topic Models for Human Emotion Analysis. Arizona State University (2015)
17. Scherer, K.R., Ekman, P.: Handbook of Methods in Nonverbal Behavior Research. Cambridge University Press, Cambridge (1982)
18. Mehrabian, A., Russell, J.A.: An Approach to Environmental Psychology. The MIT Press, Cambridge (1974)