EMOTION RECOGNITION BASED ON PRESSURE SENSOR KEYBOARDS

Hai-Rong LV, Zhong-Lin LIN, Wen-Jun YIN, Jin DONG

IBM China Research Lab

ABSTRACT

This paper describes a new approach to emotion recognition based on pressure sensor keyboards. The pressure sensor keyboard is a new product that occurs in the market recently, which produces a pressure sequence when keystroke occurs. The analysis of the pressure sequence should be a novel research area. It has been used for identity verification in our previous research. In this paper, we use the pressure sequence for emotion recognition. Three methods (global features of pressure sequences, dynamic time warping and traditional keystroke dynamics) are proposed for the emotion recognition task; then we combined the three methods together using a classifier fusion technique. Several experiments were performed on a database containing 3000 samples (from 50 individuals, including six emotions: neutral, anger, fear, happiness, sadness and surprise) and the best result were achieved utilizing all the method, obtaining an overall accuracy of 93.4%. Our technique of emotion recognition has been used for intelligent game controlling and several other applications.

***Index Terms*—**Pressure Sensor Keyboard, Emotion Recognition, Classifier Fusion

1. INTRODUCTION

Researches enabling computers to recognize human emotions have attracted increasing attention in the artificial intelligence field. It aims to make Human-Computer Interaction (HCI) more natural and friendly.

In today's HCI systems, if machines are equipped with emotion recognition techniques, they can react more appropriately, and make the interaction more natural. Other potential applications of Automatic Emotion Recognition (AER) include psychiatric diagnosis, intelligent toys, lie detection, learning environment, customer service, educational software, and detection of the emotional state in telephone call center conversations to provide feedback to an operator or a supervisor for monitoring purposes.

There have been several biometrics used for emotion recognition, such as facial expression [1], music [2] and speech signal [3]. The biometric technology employed for emotion recognition in this paper is the typing biometrics on a pressure sensor keyboard, including pressure sequences and traditional keystroke dynamics. Traditional keystroke dynamics has been widely researched in the last 30 years [4-7], whereas the keystroke pressure sequences were seldom researched in the exoteric publications until our previous research in [9]. The pressure sensor keyboard is a new product that occurs in the market recently. The most distinct difference between the pressure sensor keyboards and ordinary keyboards is that the pressure sensor keyboards can capture the pressure sequence when a key is pressed down, besides the keystroke time and the key code, whereas the ordinary keyboards can't. We can do many things with the pressure sequences produced by the pressure sensor keyboards except for user authentication, such as emotion analysis, age analysis and intelligent game controlling. In this paper, we focus our mind on the emotion recognition problem.

This paper is organized as follows. In Section 2, the database and the preprocessing methods are introduced. In Section 3, the two methods based on pressure sequences and the third one based on traditional keystroke dynamics are explained. Section 4 presents the classifier fusion method. The experiments are presented and discussed in Section 5. Finally, Section 6 presents the conclusions and future work.

2. DATABASE AND PREPROCESSING METHODS

2.1. Database

The database used in this research is collected with ourselves. The final version of the database contains 50 volunteers. Among the 50 volunteers, 50% are men, while the remaining 50% are women. First, each volunteer is asked to listen/watch carefully to a short story for each of the six emotions (neutral, anger, fear, happiness, sadness and surprise) and to immerge themselves into the situation. Once the subject is ready, he or she may read, memorize and pronounce the ten proposed utterances (one at the time), and type "how the story" using the pressure keyboard at the same time, which constitutes ten different reactions to the given situation. The volunteers are asked to put as much motion as possible, so the pressure sequence and the keystroke dynamics contains only the emotion to be elicited.

The total number of samples in our database is $10(times)\times 6(emotions)\times 50(volunteers)=3000$.

Fig.1 shows a data sample which contains n keystrokes (characters). Each sample contains n pressure sequences, n key down time T_D and n key up time T_U. Fig.2 shows six typical emotional pressure sequences. As we can see from the Fig.2, the pressure sequences of keystrokes have the characteristic that each of them has a positive impulse whereas its posterior part is negative.

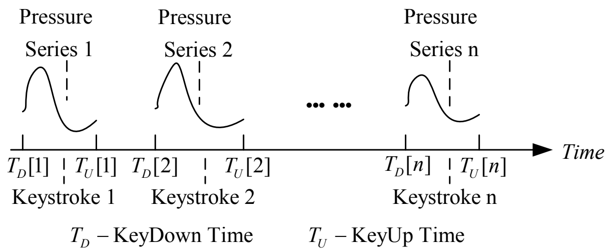

Fig.1. A sample that contains n keystrokes, which is composed of such data: n pressure sequences, n key down time T_D and n key up time T_U.

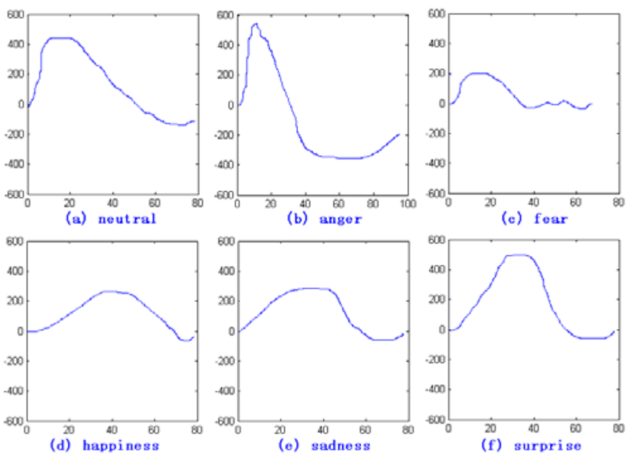

Fig.2. Six typical emotional pressure sequence samples extracted from the database. (a) neural; (b) anger; (c) fear; (d) happiness; (e) sadness; (f) surprise.

2.2. Preprocessing

The preprocessing stage contains two steps: noise removing and normalization.

2.2.1. Noise Removing

We use a average filter for noise removing with a filtering window width k. $p[i]$, i =1,2,3,…,L represent the pressure sequences, $p'[i]$ represent the sequences after noise removing.

$$p'[i] = \sum_{k=i-(w-1)/2}^{k=i+(w-1)/2} p[k] \tag{1}$$

In our experiments, k is set to 5 by experience.

2.2.2. Normalization

The purpose of normalization is to set the mean value of pressure to 0 and the standard deviation of the pressure to 1.

$$\bar{p} = \sum_{i=1}^{L} p[i] \Big/ L \tag{2}$$

$$\sigma = \sqrt{\sum_{i=1}^{L} (p[i]-\bar{p})^2 \Big/ L} \tag{3}$$

$$p'[i] = (p[i]-\bar{p})/\sigma \tag{4}$$

In equation (2)-(4), $p[i]$, i=1,2,3,…,L represent pressure sequences after noise removing; $p'[i]$ represent the sequences after normalization; $\bar{p}$ is the mean value and σ is the standard deviation.

3. OUR APPROACHES

We have presented three approaches for the emotion recognition task based on pressure sensor keyboard. The first two methods use the pressure sequences as the input and the last one uses the traditional keystroke dynamics.

3.1. Global features

We consider five global features of a pressure sequence, such as mean value ($\bar{p}$), standard deviation (σ), the difference between max and min (mm), the **P**ositive **E**nergy **C**enter (***PEC***) and the **N**egative **E**nergy **C**enter (***NEC***). Here, $\bar{p}$ and σ are computed by using equation (2) and (3) before normalization. mm , ***PEC*** and ***NEC*** are computed after normalization.

$$PEC = \sum_{i=1}^{L,p[i]>0} i \times p^2[i] \Big/ \sum_{i=1}^{L,p[i]>0} p^2[i] \tag{5}$$

$$NEC = \sum_{i=1}^{L,p[i]<0} i \times p^2[i] \Big/ \sum_{i=1}^{L,p[i]<0} p^2[i] \tag{6}$$

The five global features make up of a feature vector:

$$F = (\bar{p}, \sigma, mm, PEC, NEC)^T = (f_1, f_2, f_3, f_4, f_5) \tag{7}$$

Suppose that each sample contains n keystrokes. The global features of them are $\{F_h | h = 1,2,\ldots,n\}$.

For a new testing sample S_T, and a reference sample S_R, the distance between S_T and S_R is computed by using (8):

$$d'_G(S_T, S_R) = \frac{1}{5n} \sum_{h=1}^{n} \sum_{i=1}^{5} \left| S_T.f_{h,i} - S_R.f_{h,i} \right| \tag{8}$$

3.2. Dynamic Time Warping

Suppose that A = $a[i]$, i = 1,2,3,…,I and B = $b[j]$, j = 1,2,3,…,J represent two pressure sequences, consider the problem of eliminating timing differences between these two pressure sequences. In order to clarify the nature of time-axis fluctuation or timing differences, let's consider an i-j plane, shown in Fig.3, where sequence A and B are developed along the i-axis and j-axis respectively. The timing difference between them can be depicted by a sequence of points c = (i, j), which can be considered to represent a mapping from sequence A to B. Hereafter it's called a warping function. Then the distance between sequence A and B can be computed by (9)

$$D(A,B)=\min_F\left[\frac{\sum_{k=1}^{K}\|a(i)-b(j)\|\times w(k)}{\sum_{k=1}^{K}w(k)}\right] \quad (9)$$

Here, $w(k)$ is a nonnegative weighting coefficient.

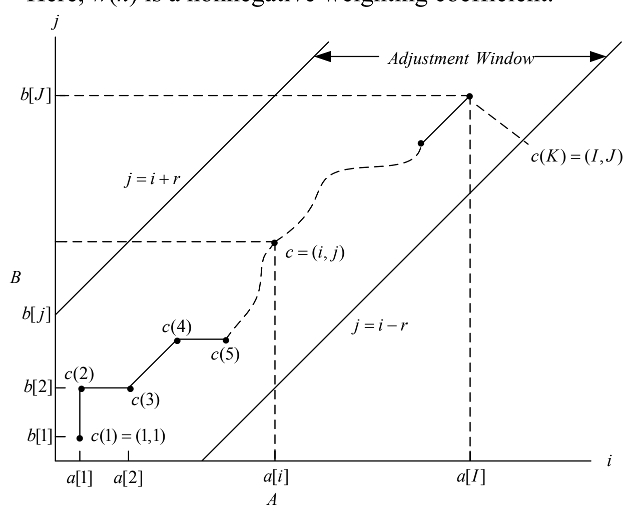

Fig.3. Warping function and adjustment window definition.

A dynamic programming technique and an adjustment window restricted on warping function are applied to compute $D(A,B)$. The detailed description was included in [8].

Suppose that there are n keystrokes in a sample for enrollment E, the pressure sequences of a sample is $S = \{P_k, k = 1,2,3,\ldots,n\}$. For a new testing sample S_T, and a reference sample S_R, the distance between S_T and S_R is computed by using (10):

$$d_D'(S_T,S_R)=\frac{1}{n}\sum_{i=1}^{n}D(P_{T,i},P_{R,i}) \quad (10)$$

3.3. Traditional Keystroke Dynamics

The original keystroke dynamics data of a sample containing n keystrokes are $T_D[i]$, $T_U[i]$ (i=1,2,3,…,n), as shown in Fig. 1. We extract two features from these data:

3.3.1. Down-Down (DD) Time

DD time is a keystroke latency defined as the time interval between successive keystrokes. This feature is represented by:

$$DD=\{dd_1,dd_2,\ldots,dd_{n-1}\} \quad (11)$$

Here,

$$dd_i=T_D[i+1]-T_D[i] \quad (12)$$

3.3.2. Up-Down (UD) Time

UD time is also a keystroke latency feature represented by:

$$UD=\{ud_1,ud_2,\ldots,ud_{n-1}\} \quad (13)$$

Here,

$$ud_i=T_D[i+1]-T_U[i] \quad (14)$$

Notice that the feature *ud* may be positive or negative according to two situations. In the first situation, key i+1 is only pressed when key i was released which results in a positive up-down time. In another situation, key i+1 is pressed while key i is still pressed, which results in a negative up-down time.

For a new testing sample S_T, and a reference sample S_R, the distance between S_T and S_R is computed by using (15):

$$d_T'(S_T,S_R)=\frac{\sum_{i=1}^{n-1}|S_T.dd_i-S_R.dd_i|+\sum_{i=1}^{n-1}|S_T.\mu d_i-S_R.\mu d_i|}{2(n-1)} \quad (15)$$

4. FUSION OF METHODS

There are three methods in our approaches. In most of the situations, the three methods would give the same decision, and sometimes different decisions are given. Though we can perform classifier fusion by voting, we select the approach based on the combination of the distance given by the three methods: $d'_G(S_T, S_R)$, $d'_D(S_T, S_R)$ and $d'_T(S_T, S_R)$. However, the three distances should be normalized. We perform the distance normalization with the method in [9] and get the corresponding $d_G(S_T, S_R)$, $d_D(S_T, S_R)$ and $d_T(S_T, S_R)$. The combination of the distances is defined by (16):

$$d_C(S,E)=w_G\times d_G(S,E)+w_D\times d_D(S,E)+w_T d_T(S,E) \quad (16)$$

Here, w_G, w_D and w_T are weights which are used to evaluate the performance of the three methods, with the summation of them to be 1. In our system, the three parameters are decided by the corresponding method's error rates, which will be described in Section 5.

5. EXPERIMENTS

As described in Section III, there are 50 volunteers in our database and it contains 60 samples for each volunteers. For the purpose of cross-validation, the 8 samples of each emotion of a volunteer are treated as training samples, the other two samples are for testing. So, 2400 samples in the training set and 600 are for testing. There totally five groups of experiments. We use nearest neighbor classifier for each volunteers.

Table 1 shows the error rates of the three methods. It's obvious that method 1 gives the best experimental results with the average error rate being 9.8%.

The last column of Table 1 shows the performance of fusion of the three methods based on the combination of the distance given by the three methods: $d_G(S, E)$, $d_D(S,E)$ and $d_T(S,E)$, by using (16). Here we use the error rates to evaluate the performance of the three methods. The weights w_G, w_D and w_T are computed by using (17). The average error rate is 6.6%, which makes 3.2% absolute improvement over method 1.

$$w_G = \frac{1}{9.8}\bigg/\left(\frac{1}{9.8}+\frac{1}{10.1}+\frac{1}{12.0}\right) = 0.3588$$
$$w_D = \frac{1}{10.1}\bigg/\left(\frac{1}{9.8}+\frac{1}{10.1}+\frac{1}{12.0}\right) = 0.3482 \qquad (17)$$
$$w_T = \frac{1}{12.0}\bigg/\left(\frac{1}{9.8}+\frac{1}{10.1}+\frac{1}{12.0}\right) = 0.2930$$

Tabel 1. Error Rates of Different Methods

Emotion	Method 1	Method 2	Method 3	Combined
Neutral	7.2	7.8	12.2	5.8
Anger	8.6	8.4	12.8	6.6
Fear	7.8	7.6	12.0	6.4
Happiness	**14.4**	**14.8**	**12.4**	8.2
Sadness	**14.2**	**14.4**	**10.6**	8.4
Surprise	6.4	7.4	12.2	4.4
Average	**9.8**	**10.1**	**12.0**	**6.6**

We can also find from Table 1 that, pressure series (method 1 and method 2) are not good enough to distinguish happiness and sadness, but the traditional keystroke dynamics (method 3) can. It helps to improve the overall recognition accuracy of the combined classifier.

6. CONCLUSION

This paper presents a new methodology for emotion recognition through typing biometrics features produced by Pressure Sensor Keyboards. The use of pressure sequences and traditional keystroke dynamics for emotion recognition is novel. Some experiments were conducted and the best performance was achieved when we combined the three methods (global features, dynamic time warping and traditional keystroke dynamics), obtaining an average error rate of 6.6%. It proves that typing biometrics features is a feasible way for emotion recognition.

However, we only use simple but effective feature extraction methods for pressure sequences. As the use of pressures sequence for emotion recognition is very novel, we'll test more models and approaches to solve the problem in the future.

8. REFERENCES

[1] K.S. Byun, C.H. Park, K.B. Sim, "Emotion recognition from facial expression using hybrid-feature extraction," *ICE Annual Conference*, 2004

[2] Y.H. Yang, Y.C. Lin, Y.F. Su, H.H. Chen, "Music emotion classification: a regression approach," *IEEE International Conference on Multimedia and Expo*, 2007

[3] J. Liu, C. Chen, J.J. Bu, M.Y. You, "Speech emotion recognition using an enhanced co-training algorithm," *IEEE International Conference on Multimedia and Expo*, 2007

[4] L. C. F. Araujo, L. H. R. Jr. Sucupira, "User authentication through typing biometrics features," *IEEE Trans. Signal Processing*, 53(2): 851-855, 2005

[5] F. Monrose, M. K. Reiter, "Password hardening based on keystroke dynamics," *in Proc. 6th ACM Conf. Computer Security*, Singapore, Nov. 1999

[6] W.G. Ru, J.H.P. Eloff, "Enhanced password authentication through fuzzy logic," *IEEE Expert*, 17(6): 38-45, 1997

[7] S. Bleha, C. Slivinsky, B. Hussain, "Computer-access security systems using keystroke dynamics," *IEEE Trans. Pattern Anal. Machine Intell.*, 12(12): 1217-1222, 1990

[8] H. Sakoe, S. Chiba, "Dynamic programming algorithm optimization for spoken word recognition," *IEEE Trans. Acoust., Speech, Signal Processing*, 26(1): 43-49, 1978

[9] H.R. Lv, "Biologic verification based on pressure sensor keyboards and classifier fusion techniques," *IEEE Trans. on Consumer Electronics*, 52(3): 1057-1063, 2006