Sensors & Transducers, Vol. 163 , Issue 1, January 2014, pp. 316-322

Sensors & Transducers

Research on Interaction-oriented Gesture Recognition

[1] Lu Huang, [1] Zongkui Zhou, [2] Rong Liu
[1] Key Laboratory of Adolescent Cyberpsychology and Behavior(CCNU), Ministry of Education, and Key Laboratory of Human Development and Mental Health of Hubei Province,
School of Psychology, Central China Normal University,
Wuhan, Hubei, 430079, P. R. China
[2] College of Physical Science and Technology, Central China Normal University,
Wuhan, Hubei, 430079, P. R. China
[1] Tel.: +86-027-67868632, fax: +86-027-67868632
E-mail: bonnie.huang@mails.ccnu.edu.cn, zhouzk@mail.ccnu.edu.cn

Received: 23 December 2013 /Accepted: 20 January 2014 /Published: 31 January 2014

Abstract: This thesis designs a series of gesture interaction with the features of the natural human-machine interaction; besides, it utilizes the 3D acceleration sensors as interactive input. Afterwards, it builds the Discrete Hidden Markov Model to make gesture recognition by introducing the collection proposal of gesture interaction based on the acceleration sensors and pre-handling the gesture acceleration signal obtained in the collection. In the end, the thesis proofs the design proposal workable and effective according to the experiments. *Copyright © 2014 IFSA Publishing, S. L.*

Keywords: Natural human-machine interaction, Gesture recognition, Acceleration sensors, Discrete hidden Markov model.

1. Introduction

With the constant improvement of human civilization and living standard as well as the development of computer science, human-computer (man/machine) interaction technology steps into the spotlight in the information industry. People expect that the future man-machine interaction will be as natural, accurate and fluent it is done between humans. In this regard, natural man-machine interaction is a way in which computer is able to accurately recognize, understand and act the content, idea and intention expressed by the other side of the interaction naturally and friendly.

As a component of body language, gestures express people's emotion, attitude and intention in a natural and straightforward way. People could manipulate any concrete object with their hands by vision, touch and force sense. Hands of human beings have become the media connecting human beings and the world. In this way, hands have become the most direct, natural and practical medium between human and external world [1].

With the development and popularization of mobile computer technology, gesture interaction based on wearable equipment and device becomes a hot spot in recent man-machine interactive research. Now most gesture recognition research is based on graphic processing – a way which is of highly restriction and dependency on external environment – therefore breakthrough is nearly impossible. This paper is going to conduct research on gesture recognition based on 3D acceleration sensors by introducing correlated knowledge of man-machine interactive model.

Article number P_RP_0069

2. Gesture Recognition Based on Natural Human-Machine Interaction

From Hermann Hollerith's mechanical tabulator based on punched cards which can rapidly tabulate statistics from millions of pieces of census data, to the man-machine interaction based on computer for the technicians in the last century, and to the emerging of mouse and WIMP interface, man-machine interaction technology and its application have changed people's life regularly and brought significant convenience to the world.

As the computer and the Internet dashing forward, people began to realize that the existing was of interaction was still far from the state of perfectness due to the monotonousness of interactive content and mode, as well as its high dependency on users and environment. Amore affluent and diversified man-machine interactive mode was the key to change the status quo.

The appearing of the so-called somatic game (or motion sensing game) was a proof of natural man-machine interaction's advantages: natural, user-friendly and highly effective. The natural man-machine interaction aims at a more direct and human-friendly way of interaction between users and computers and making the machines to study human's way of thinking so as to complete the interactive task following human's conventional habit through the multiple-channel approach including gestures, voice and facial expressions [2].

Nintendo's Wii has accelerometer sensor, gyroscope and infrared ray device in its joysticks (accelerometer sensor proves data of speed and displacement, gyroscope provides data of azimuth and infrared ray device is used for positioning). By using the joystick, gamers can get full control of direction and speed and other manipulation in games through natural and direct body actions. J. Baek and I. Jang use 2D accelerometer sensors to collect data of user's consecutive motion and poses, and decide the motion status of user holding a mobile phone via the change of acceleration in two axes [3]. The acceleration sensor based recognition is less likely to be influenced by external environment and can even be handled by one hand. Hou Xiangfeng and his team collect data of tester's pace and pose through an accelerometer sensor attached to the tester's waist, and conclude user's pose and gesture via the periodical acceleration change of the three axes [4]. So in the model of portable and mobile computers, gesture collection based on image or graphic processing cannot be widely applied subject to the size of the device and external circumstances (source of light, background etc.) whereas gesture collection based on accelerometer sensor is able to recognize gestures effectively.

Perceived from the angle of technical implementation, when the interactive mode is established, the core of the mission is to design appropriate gesture and collecting equipment in order to "understand" what the users mean by their gestures.

3. Gesture Acceleration Collection

The Nintendo-like natural man-machine interactive device began to lead the trend in the manufacturing industry with the arising of Wii in 2006. After that, Apple and Nokia followed suit and brought sensors into mobile phones which created new interactive experience for the customers, the feedback was good.

In everyday life, people are more likely to use arms with dynamic and directional gestures instead of gestures of fingers and static poses. Wrist is one of the main force-generating points of arm which is proved to have biological characteristics in 3D movement. Just as wearing a watch or other arm-attached things, a wearable collecting device could cause little inconvenience to the users when collecting the 3 axis acceleration so that they can focus on natural and free interaction. The 3 axis accelerometer sensor can collect data like the observed object's state and track of movement.

Gestures adopted in man-machine interaction are conscious and intentional. For the design of natural gestures, they must be of this kind: easy to recognize, easy to use, direct and straightforward. Assume that there is a set of motions which includes 8 separate ones: up, down, left, right, left circle, right circle, move forward and move backward. The gestures in collection are depicted in Fig. 1 in which the black line represents the track of the hand.

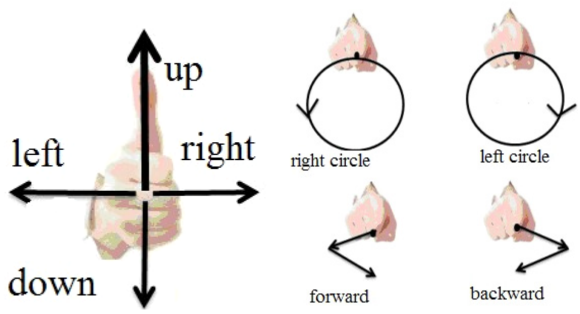

Fig. 1. Gestures.

3.1. Hardware

The gesture collection system is made up of two components: the accelerometer sensor MMA7260 which is responsible for the collection of the wrist's 3D acceleration signal, and micro controller ADUC841 which converts the analog signal of 3 axis acceleration captured by MMA7260 to digital signal and conduct real-time transition to PC through serial port. Real-time transition, small size, low energy cost and highly compatible are the most prominent advantages of this device.

3.2. Data Collection of Gesture Acceleration

During the collection process, acceleration data of three directions is shown on the man-machine interaction interface. Fig. 2 shows the real-time data transition on the PC interface when it is connected with the collecting system. The first row of the data analysis column represents the count of sample points while the third, fourth and fifth rows represent corresponding acceleration data of axis X, Y and Z. The collection and transition are done by the PC. Apart from that, the parameters of serial port and data collection in the experiment can be configured, recorded and saved.

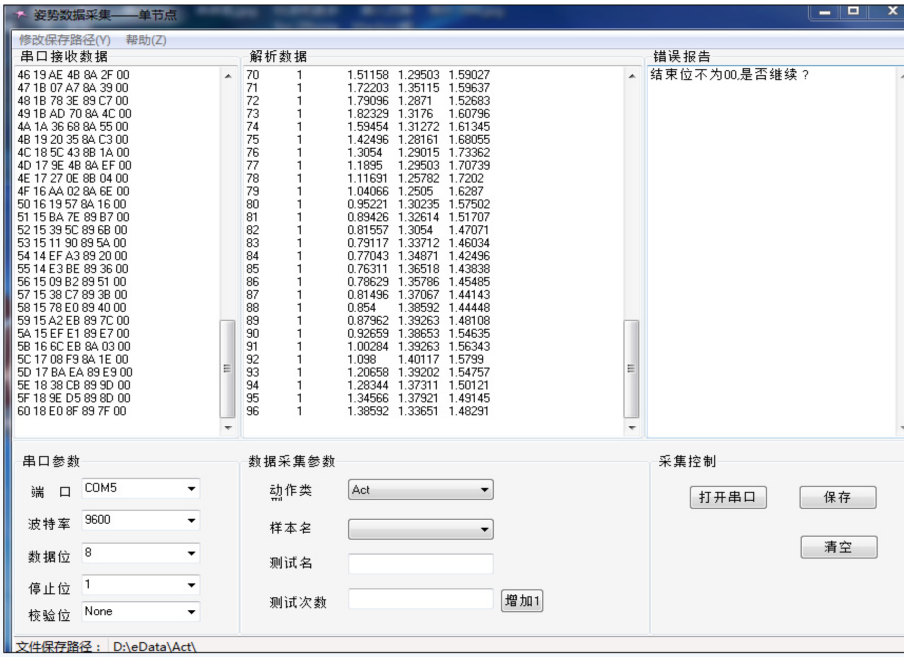

Fig. 2. Interface of collecting system PC.

4. The Framework of Recognition System

Fig. 3 shows the framework of gesture recognition process. It is comprised of two parts: the model training and recognition of gestures, the data necessary for each one must go through data processing by some programs. For model training, actually it is a model optimization process according to the data of training set after the initialization of model parameters; and recognition is to match these optimized gestures with corresponding gesture data.

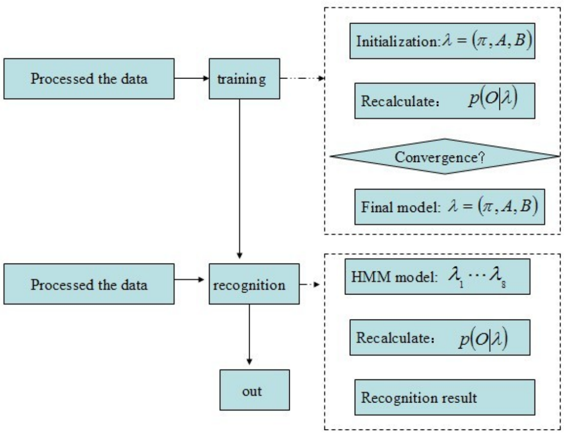

Fig. 3. Framework.

4.1. Data Preprocessing and Feature Extraction

Before gesture recognition, simple preprocessing of the collected data need to be done in order to enhance recognition rate and simplify the recognition complexity. By using normalization the range of the data is confined between [-1, 1] with the smooth process of the data waveform. For an example, the before and after preprocessing sample data waveform can be seen in Fig. 4(a) and 4(b) respectively.

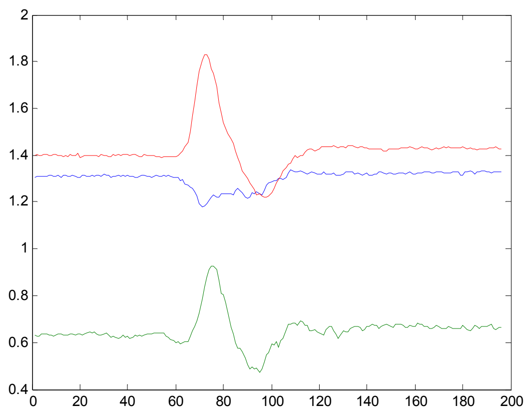

Fig. 4 (a). Before preprocessing sample data waveform.

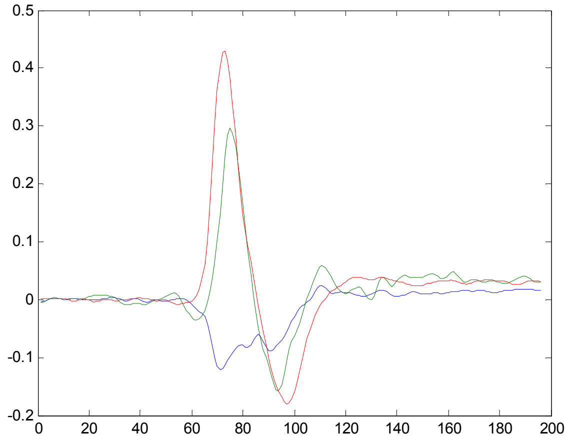

Fig. 4 (b). After preprocessing sample data waveform.

Since the duration of gesture acceleration signal in this experiment is comparatively short, a re-sampling in five times sample rate (125 HZ) is conducted to extend the length of the sample signal. A slide window with frame length set to 32 sample points and frame offset 10 sample points implements the framing work of the re-sampled sample. After the framing work is done, the mean value of acceleration signal of every frame is computed as a discrimination feature (the mean value here refers to mean value of sample point amplitude in every frame which stands for change in gesture of the sampled object). The choice and extraction of features determine the performance of classifier and degree of recognition rate. Because the feature of the extracted mean value is a set of vectors consisting of floating numbers, it requires huge storage space of the computer and the computing efficiency is low. In purpose of simplifying computing and easing the burden of CPU, K-means algorithm needs to be introduced to discretize the vectors.

4.2. Hidden Markov Model

Gestures can be divided into a series of time-varying "meta gestures" (the so called meta-gesture is the smallest unit [5] of a gesture which can be described by Hidden Markov Model (HMM)). The Discrete Hidden Markov Model can be expressed as:

$$\lambda = (\pi, A, B), \tag{1}$$

A, B and π configure the initial value. A is a state transition matrix whose spontaneous transition probability is set to 0.75 and its probability of transferring to next state is 0.25. When the state value s=3, it is a state transition matrix.

$$A = \begin{bmatrix} 0.75 & 0.25 & 0 \\ 0 & 0.75 & 0.25 \\ 0 & 0 & 1 \end{bmatrix} \tag{2}$$

Now the corresponding parameter π is:

$$\pi = (1,0,0), \sum_{i=1}^{S} \pi_i = 1 \tag{3}$$

B represents the probability distribution of j, and the model sets off "C" observable symbols, that is, the number of clusters mentioned in the feature extraction. The initial value of B is estimated by using method of mean value.

The optimization of model is to re-estimate each parameter of it till these parameters are set to some given condition: eps convergent or the maximum iteration number amounts to 10.

$$eps = \log p(O|\lambda) - \log P(O|\lambda_0) \tag{1}$$

4.3. Gesture Recognition

A complete gesture algorithm includes: Effective extraction of feature data in collected data. The construction of models, basically, every gesture corresponds with a HMM model. Since there are 8 gestures in this paper, there has to be 8 HMM models. First step is model parameter initialization, then training of each model. After that the optimal model is chosen in order to describe each sample best.

The classification of gestures, putting the sample of gesture data into 8 corresponding HMM model in order to get 8 results of probabilities, the gesture of the maximum probability is the sample's recognized gesture category.

5. Experimental Result Analysis

In this experiment, 2400 samples of 8 gestures (collect for 30 times each gesture each subject) are extracted from 10 examinees (5 men and 5 women between ages 20 to 28).

K-fold Cross Validation method is adopted for cross validation in the experiment, K=10. For the measure of classifier performance, it is "average recognition rate ± standard deviation" in which the standard deviation is the discretization degree of validated recognition rate. And then confusion matrix is introduced to analyze HMM's discrimination level among different samples. The results of status numbers (11 status numbers from status 2 to 12) of particular and non-particular people are also compared.

Table 1 shows the highest recognition rate, lowest recognition rate, average rate and standard deviation of particular people in 11 statuses, and Table 2 shows the highest recognition rate, lowest recognition rate, average rate and standard deviation of non-particular people in 11 statuses respectively.

From Fig. 5, it is obvious that when the status number is 4, the average recognition rate of particular people increases fastest. In this circumstance, from Table 3, the lowest recognition rate is 95 % while the highest is 100 %; for the whole picture (Fig. 5), the average recognition rate 98.79 % under status number 9 is the highest with the lowest rate 95.34 % and the highest 100 %.

Take a look at Table 5, from the recognition rate and recognition rate curve (Fig. 6) we can see that when the status number is 4, the escalating rate is the highest for non-particular people: the lowest recognition rate is 85.34 % while the highest is 100 %. For the whole picture, the average recognition rate 97.21 % is highest when the status number is 11 with the lowest rate 91 % and the highest 100 %.

Table 1. The gesture recognition rates of particular people in different statuses based on DHMM.

Status number	Average rate	Highest recognition rate	Lowest Recognition rate	Standard deviation
2	0.9742	1	0.8792	0.034683
3	0.9775	1	0.8958	0.031007
4	0.9808	1	0.9083	0.027372
5	0.9821	1	0.9000	0.030178
6	0.9817	1	0.9125	0.027089
7	0.9846	1	0.9083	0.027642
8	0.9850	1	0.9250	0.023750
9	0.9879	1	0.9167	0.025493
10	0.9871	1	0.9208	0.024410
11	0.9850	1	0.9250	0.022498
12	0.9858	1	0.9208	0.024072

Table 2. The gesture recognition rates of non-particular people in different statuses based on DHMM*.

Status number	Average rate	Highest recognition rate	Lowest Recognition rate	Standard deviation
2	0.9083	0.9208	0.8917	0.008562
3	0.9342	0.9417	0.9208	0.006747
4	0.9538	0.9583	0.9417	0.005361
5	0.9520	0.9583	0.9417	0.005974
6	0.9571	0.9667	0.9417	0.009633
7	0.9558	0.9667	0.9375	0.009461
8	0.9629	0.9792	0.9417	0.012797
9	0.9629	0.9708	0.9542	0.006038
10	0.9658	0.9750	0.9583	0.007027
11	0.9721	0.9792	0.9667	0.004414
12	0.9713	0.9833	0.9542	0.009909

* Non-particular people mean to put all the examinees' data together and choose the data from one of them as test set [6].

Table 3. The confusion matrix of particular people for status number 4.

Gesture \ Recognition	Back	Down	Forward	Left	Left circle	Right	Right circle	Up	Possibility
back	300	0	0	0	0	0	0	0	1
down	12	285	0	0	0	3	0	0	0.95
forward	7	0	292	0	0	1	0	0	0.9734
left	0	0	2	292	0	3	3	0	0.9734
left circle	0	0	1	0	299	0	0	0	0.9967
right	1	0	6	0	0	290	3	0	0.9667
right circle	0	0	0	0	0	0	300	0	1
up	0	0	3	0	0	0	1	296	0.9867

Table 4. The confusion matrix of particular people for status number 9.

Gesture \ Recognition	Back	Down	Forward	Left	Left circle	Right	Right circle	Up	Possibility
back	300	0	0	0	0	0	0	0	1
down	10	286	1	0	0	2	0	1	0.9534
forward	3	0	296	0	0	1	0	0	0.9867
left	0	0	0	300	0	0	0	0	1
left circle	0	0	1	0	299	0	0	0	0.9967
right	0	0	6	0	0	293	0	1	0.9767
right circle	0	0	0	0	0	0	300	0	1
up	0	0	0	3	0	0	0	297	0.99

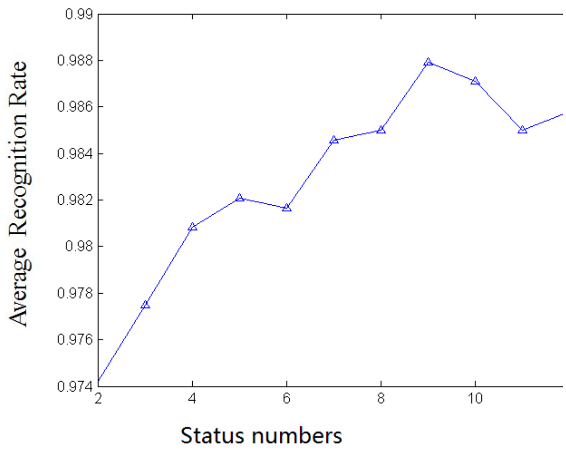

Fig. 5. Average recognition rate curve of particular people for 11 status numbers.

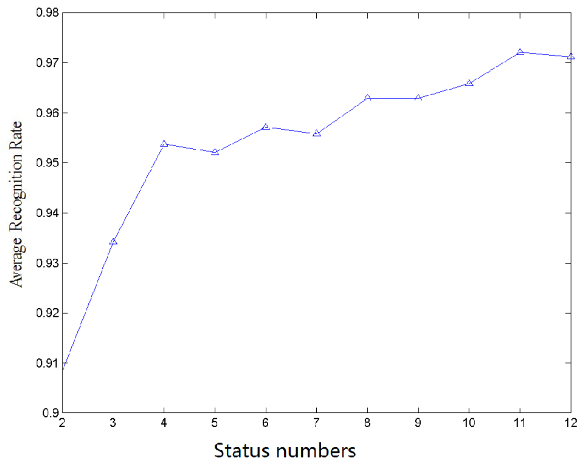

Fig. 6. Average recognition rate curve of non-particular people for 11 status numbers.

Table 5. The confusion matrix of non-particular people for status number 4.

Gesture \ Recognition	Back	Down	Forward	Left	Left circle	Right	Right circle	Up	Possibility
back	300	0	0	0	0	0	0	0	1
down	39	256	5	0	0	0	0	0	0.8534
forward	0	4	256	0	20	9	11	0	0.8534
left	0	0	0	300	0	0	0	0	1
left circle	0	0	0	0	300	0	0	0	1
right	0	0	0	10	0	290	0	0	0.9667
right circle	0	0	0	0	0	3	297	0	0.99
up	0	0	10	0	0	0	0	290	0.9697

Table 6. The confusion matrix of non-particular people for status number 11.

Gesture\ Recognition	Back	Down	Forward	Left	Left circle	Right	Right circle	Up	Possibility
back	296	0	4	0	0	0	0	0	0.9867
down	10	288	2	0	0	0	0	0	0.96
forward	0	0	273	0	0	9	18	0	0.91
left	0	0	0	300	0	0	0	0	1
left circle	0	0	0	0	300	0	0	0	1
right	0	1	1	0	0	296	2	0	0.9867
right circle	0	0	0	0	0	8	292	0	0.9734
up	4	0	0	0	0	8	0	288	0.96

6. Conclusions

“People-oriented” spirit leads the way in this paper to create a natural man-machine interaction system. By adopting the new way of man-machine interaction of non-precision input, and the equipment of the system is low-cost, small-sized, low energy cost, real-time and portable as well. The experiment data demonstrates that this system can achieve high recognition rate while providing the users with friendly and effective man-machine interaction.

Acknowledgements

We gratefully acknowledge financial support from the major program of the National Social Science Foundation (No. 11&ZD151); and the national "Twelfth Five" science and technology support program (No. 2011BAK08B05).

References

[1]. Ren Hai-Bing, Zhu Yuan-Xin, Xu Guang-You, Zhang Xiao-Ping, Lin Xue-Yin, Hand gesture segmentation and recognition with complex backgrounds, *Acta Automatica Sinica*, Vol. 28, No. 2, 2002, pp. 256–261.

[2]. Beijing Human: Human-Computer Interaction in the Year 2020, (http://research.microsoft.com/en-us/um/Cambridge/projects/hci2020/download.html).

[3]. Jonghun Baek, Ik-Jin Jang, Kee Hyun Park, Hyun-Soo Kang, Byoung-Ju Yun, Computer interaction for the accelerometer-based mobile game, in

Proceedings of the International Conference on Embedded and Ubiquitous Computing (EUC 2006), Seoul, Korea, 15-19 August 2006, *Lecture Notes in Computer Science*, Vol. 4096, pp. 509-518.

[4]. H. Xiangfeng, L. Rong, Z. Zhaofeng, Application of accelerometer MMA7260 in the gait feature extraction, *Chinese Journal of Sensors and Actuators*, Vol. 20, Issue 3, 2007, pp. 507-551.

[5]. X. Chonggao, "Met-motor" and Meta-cognition on assimilate models of motor learning process, *Journal of XiAn Institute of Physical Education*, Vol. 18, Issue 1, 2001, pp. 102-104.

[6]. J. Feng, G. Wen, W. Chunli, Y. Hongxun, Z. Debing, Development in signer-independent sign language recognition and the ideas of solving some key problems, *Journal of Software*, Vol. 18, Issue 1, 2007, pp. 477-489.
