Ping Pong Motion Recognition based on Smart Watch

Zengjun Fu [a], Kuang-I Shu [b] and Heng Zhang [c, *]

School of Computer and Information Science, Southwest University, Chongqing, 400715, China.

[a]fuzengjun123@163.com, [b]kuangishu@swu.edu.cn, [c, *] dahaizhangheng@163.com

Abstract. Smart watches have become one of the most representative devices in wearable devices because of their unique advantages such as integration, portability, reliability, stability, universality and low environmental dependence. At present, it is mainly used for the monitoring of health indicators such as human heart rate. Whole-body inertial sensing devices cannot meet the actual needs of the general public for virtual sports because of high prices and inconvenient wear. In this paper, a single piece smart watch is used to study the recognition of the most common actions in table tennis which is a kind of fast-moving sport and has many fans through an improved convolution neural network model. The final experimental results show that the recognition accuracy reaches 95.46%, which can basically meet the needs of amateurs' motionSports.

Keywords: Smart Watch; Inertial perception; Convolutional neural network; Motion recognition.

1. Introduction

Ping-pong sports, as a kind of enthusiastic ball game, is popular in some countries because of its simple site requirements. However, table tennis requires long-term effective training to achieve sports stereotypes, so as to better control the table tennis with the body. At present, only a handful of professional athletes can master these skills. It is difficult for ordinary people (even some ordinary athletes) to master it. This is not conducive to the promotion of the sport as a global project. In addition, virtual table tennis is not only more interesting than the actual table tennis in some aspects, but also can become a popular platform for real-time communication. The accurate recognition of the important movements of the table tennis in a convenient situation is the key to achieving the popularity of its auxiliary training and the somatosensory games of it.

In current research on motion recognition, there are two main directions: vision based [1-2] and sensor based [3-4]. Vision-based motion recognition technology has many limitations and shortcomings: a) High requirements on the accuracy of the equipment and the fixed position of the equipment cannot be used flexibly, and it is affected by light; b) The scope of perception is limited. It needs to be detected at a specific angle and within a certain distance. Inertial sensors, especially integrated wearable micro inertial sensors, have the advantages of small size, high precision, low energy consumption, and low environmental dependence, which can effectively compensate for the lack of visual recognition technology. It has been widely used in various fields such as competitive sports [5-6], rehabilitation treatment [7-8] and somatosensory games [9]. In this paper, we recognize the main ping-pong movements through nine-axis inertial sensors (accelerometers, gyroscopes, and magnetic field sensors) integrated in smart watches, which is universal and convenient and can achieve high recognition accuracy.

In order to achieve a more harmonious computing environment, an effective computing model is very necessary. Article [10] presents a tensor-based cloud-edge computing framework for providing CPSS services. In the field of behavior recognition, many scholars have done relevant research, such as the article [11] proposed an algorithm which is based on a numerical statistical analysis technique called n-mode analysis. As an efficient deep learning model, Convolutional Neural Network (CNN) can effectively extract the depth features of data and it has achieved good results in many aspects [12-13]. Many researchers have already proposed improved algorithms for CNN [14-15]. We try to recognize the ping-pong action through a specially designed dropout convolutional neural network model and make some analysis of the results. The experiments show that it has a good effect on inertial sensor-based action recognition. The experimental process is shown in Figure 1.

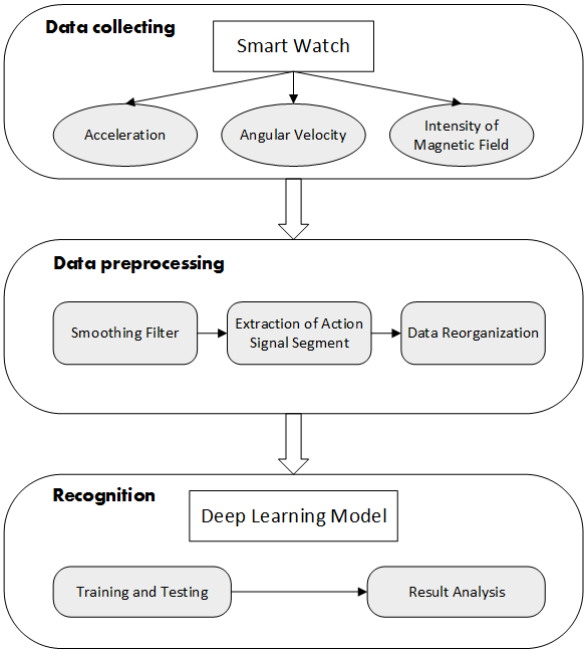

Fig. 1 The process of ping-pong action recognition

2. Data Collecting and Processing

2.1 Data Collecting

The data acquisition device we used is HUAWEI WATCH 2 smart watch, which has various sensors such as acceleration, gyro, magnetic field sensing, heart rate, pneumatic pressure, CAP capacitance, ALS/ambient light sensing, and positioning system. Acceleration, angular velocity and magnetic field intensity are used in this paper, and their corresponding units are m/s^2, $Radian/s$, and μT. For each sensor, we obtain data in the direction of the three axes of X, Y, and Z, and the directions of the three sensor axes are the same.

The process is as follows: First, we developed an android wear application to acquire real-time data of acceleration, angular velocity and magnetic field strength on smart watch at a frequency of 50 Hz, and we transmitted them to a mobile phone via Bluetooth. Then, the data is received through an android application on the mobile phone, and at the same time they are forwarded to a PC via WiFi. Finally, the data is received and stored on the PC by a specific Java server program. As shown in Figure 2.

Fig. 2 Data Collecting

In the experiment, we asked the participants to wear a smart watch and complete eight basic actions of table tennis, and we trained all participants before the experiment. The actions to be recognized are: Forehand Attack, Forehand Drive, Forehand Chop, Forehand Pick, Backhand Dial, Backhand Drive, Backhand Chop, Backhand Twist. We collected data from 12 college volunteers aged 18-28, including 6 males and 6 females. A total of 2,275 valid samples were collected, of which 1,147 were male and 1,128 were female. Each action and the corresponding number of samples are shown in Table 1.

2.2 Action Signal Detection

The first step of recognition is to segment the action signal accurately. In the period of no motion, the signal is stable and has a small variance, and within the action interval, the signal fluctuates greatly

and has a large local variance. Therefore, we can set a sliding time window w to detect and divide the action signal data by controlling the variance within the window. The specific process is: First, we calculate the variance of each axis for each sensor data in the window, and then compare the sum of each axis's variance for each sensor with the set threshold. When they satisfy the constraint at the same time, we judge it as the window containing the action signal segment, where the overlap size is $w/3$.

After preprocessing the data, we labeled 2275 action data with eight corresponding actions (as shown in Table 1). We integrate the actions with the corresponding labels, and store them according to different actions, different individuals, and different organizational forms used by the model. Finally, the corresponding data files are called separately during the deep learning model training.

Table 1. Action name and corresponding label number

Action Name of Table Tennis	label number
Forehand Attack	1
Forehand Drive	2
Forehand Chop	3
Forehand Pick	4
Backhand Dial	5
Backhand Drive	6
Backhand Chop	7
Backhand Twist	8

3. Recognition Model

There are many methods of action recognition, such as decision trees, Bayesian methods, Nearest Neighbor, Support Vector Machines, and Neural Networks [16], et al. Relevant research shows that the recognition method based on deep learning has a good effect [17-18], and they have a wide range of applications in the field of computer vision. In this paper, the convolutional neural network model is used to identify and analyze the ping-pong action based on inertial sensing data.

The convolutional neural network can automatically learn data features through multi-layer non-linear transformations, and has strong expressive ability and learning ability. In article [19-22], convolution neural network is used to extract features. CNN has the characteristics of local connection, weight sharing and pooling operation, which can effectively reduce the network complexity and make it easy to train and optimize. It is widely used in image recognition and speech recognition. The structure of convolutional neural network used in this paper is shown in Figure 3.

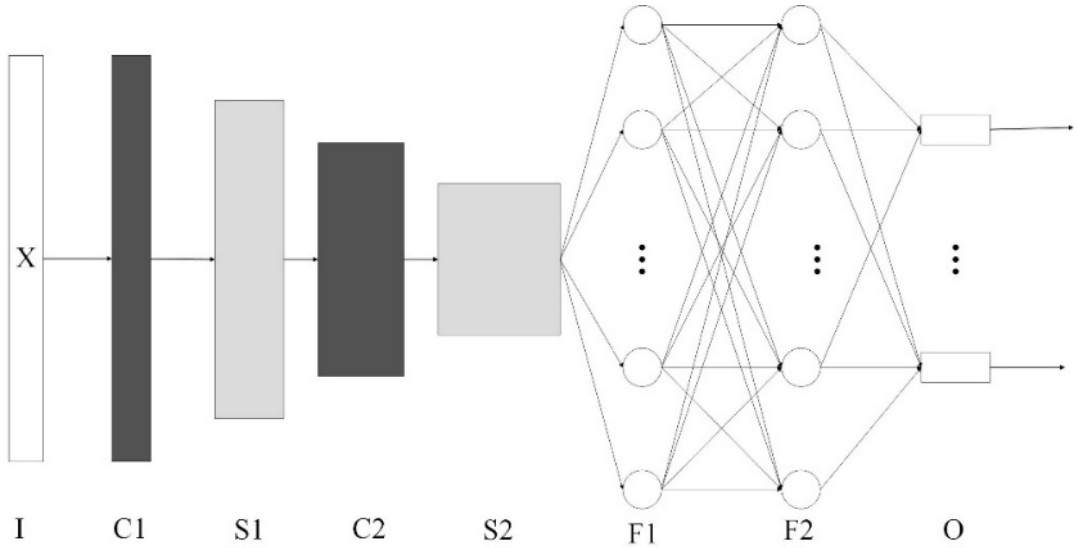

Fig. 3 Network structure.

where C1 and C2 are convolutional layers; S1 and S2 are down sampling layers; F1 and F2 are fully connected layers.

Input layer is the original data of the action signal segment $X = (\boldsymbol{x_1}, \boldsymbol{x_2}, \ldots, \boldsymbol{x_n})$; n is the number of input samples. The formula for convolutional layer is

$$x_j^l = f\left(\sum_{i \in M_j} x_i^{l-1} * k_{ij}^l + b_j^l\right) \tag{1}$$

where x_j^l represents the j_{th} feature map of the l_{th} layer; M_j is the set of input feature maps; k_{ij}^l is the j_{th} convolution kernel of layer l; b_j^l is the bias; $f(\cdot)$is the activation function. In this paper, ReLU (Rectified linear unit) [23] function is used as the activation function (Formula 2).

$$f(x) = \max(0, x) \tag{2}$$

Down sampling layer follows the convolutional layer and corresponds to the feature map in previous layer, with spatially invariant features [24]. The formula is

$$x_j^l = f(w_j^l down(x_j^{l-1}) + b_j^l) \tag{3}$$

where, $w_j^l$is the weight; $b_j^l$is the bias; $down(\cdot)$is the downsampling function.

After two convolutional-pooling layers, two fully connected layers are connected, each neuron of which is connected to all neurons in previous layer. The fully connected layer can integrate the local information with class discrimination among convolutional layer or pooling layer [25]. The activation function is still the ReLU function. Finally, at output layer, the sample is classified by a SoftMax function.

We provide the processed raw data described in Section 3 directly to the model for training, which does not require manually extract the features of the signal. The experimental results show that for the three kinds of inertial sensing data used in this paper, CNN can effectively extract the motion signal features and achieve high recognition accuracy.

4. Experiments and Analysis

In order to explore the recognition effect of ping-pong action when the sensors are combined, in the case of a single sensor, two sensor combinations, and three sensors used simultaneously, we performed a series of comparative experiments. We randomly selected 475 samples for testing and the remaining 1800 samples for training. In the experiment, we set the batch size to 50 and the number of iterations to 200.

4.1 Recognition of the Three Sensors

(1) A Single Sensor

First of all, we separately train and test the triaxial data of accelerometer, gyroscope and magnetometer. The accuracy of recognition varies with the number of iterations as shown in Figure 4.

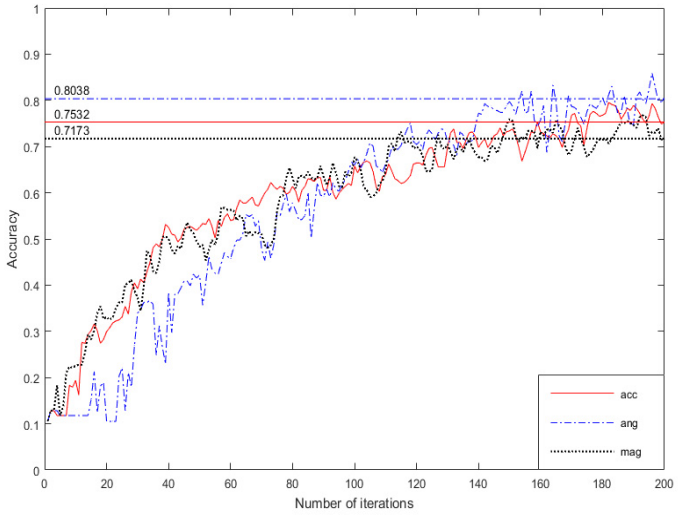

Fig. 4 Experimenting with a single sensor

As can be seen from Fig. 9, in the process of training for using three sensor data separately, the recognition accuracy rate becomes higher and higher with the number of iterations and tends to be stable after 200 iterations. The accuracy of accelerometer, gyroscope and magnetometer reached 75.32%, 80.38% and 71.73% finally.

(2)Two Combined Sensors

Next, we explored the recognition effect when combining two sensors. We combine two of the accelerometer data, gyroscope data, and magnetic field sensor data, and then train and test them. The accuracy of recognition varies with the number of iterations as shown in Figure 5.

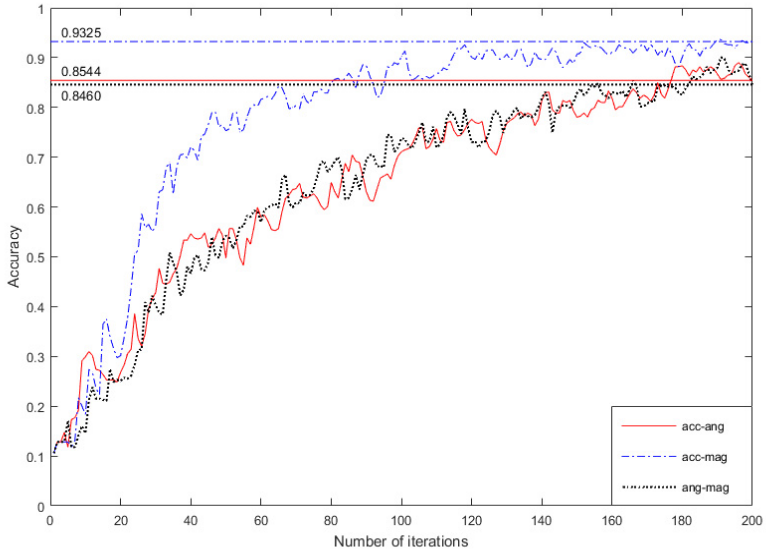

Fig. 5 Experimenting with two combined sensors

As can be seen from Fig. 10, during the training process, the combination of acceleration-magnetic field works best, achieving an accuracy of 93.25% after 200 iterations. The accuracy of other two combinations (acceleration-angular velocity, angular velocity-magnetic field) reached 85.44% and 84.60% respectively. Overall, the combined recognition of the two sensors is superior to that of a single sensor.

(3) Recognition of the Combination of Three Sensors

Finally, we experimented with the comprehensive utilization of the three-sensor data, and finally reached an accuracy rate of 96.62% (Figure 6). It can be seen that the combined use of the three sensors has the best recognition effect compared to the case where a single sensor and the combination of two sensors.

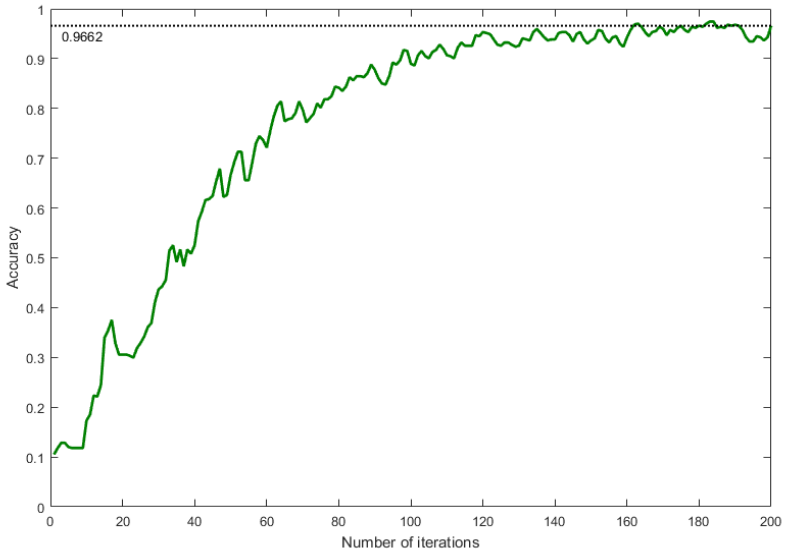

Fig. 6 Experimenting with the combination of three sensors

4.2 Comprehensive Comparison

In order to avoid the occasionality during the experiment, we did twelve experiments for each case of 5.1-5.3, so that we can observe the experimental results more accurately and objectively. The accuracy of recognition for all experiments are shown in Table 2

Table 2. The results of twelve experiments

Experiment number	Acc	Ang	Mag	Acc_ang	Acc_mag	Ang_mag	Acc_ang_mag
1	0.7806	0.8059	0.7658	0.8924	0.9367	0.8797	0.9726
2	0.7722	0.8312	0.7975	0.8565	0.9515	0.8523	0.9367
3	0.7384	0.7785	0.8186	0.8565	0.8966	0.8460	0.9409
4	0.7532	0.7743	0.7785	0.9093	0.8987	0.8713	0.9515
5	0.7679	0.8270	0.7954	0.8671	0.9241	0.8249	0.9662
6	0.7574	0.7932	0.8101	0.8544	0.9008	0.8671	0.9620
7	0.7447	0.7869	0.7890	0.8797	0.9156	0.8861	0.9473
8	0.7869	0.7785	0.7173	0.8671	0.9072	0.8734	0.9620
9	0.7574	0.8038	0.7321	0.8924	0.8924	0.8544	0.9451
10	0.7637	0.8228	0.7869	0.8776	0.9325	0.8629	0.9599
11	0.7511	0.8207	0.7869	0.8397	0.9177	0.8840	0.9599
12	0.7468	0.8207	0.7342	0.8987	0.9325	0.8565	0.9515

It can be seen that the effect of recognition is different for different sensor combinations and different ways of data organization. Even using the same sensor and under the same data-organization way, the results are not the same when we repeat the experiment. For more objective, more accurate and more direct comparison result, after removing the maximum and minimum values, we calculated the average accuracy in each case and listed them in Table 3.

Table 3. Average recognition accuracy

Sensor	Acc	Ang	Mag	Acc_ang	Acc_mag	Ang_mag	Acc_ang_mag
Accuracy	0.7595	0.8038	0.7776	0.8742	0.9162	0.8648	0.9546

In the case of a single sensor, the recognition accuracy of the gyroscope data is slightly higher than the other two sensors. In the combination of two sensors, the combination of acceleration data and magnetic field data is the best, achieving an accuracy of 86.48%. In all cases, the best is the comprehensive use of three sensors, including acceleration, angular velocity and geomagnetic sensing, with a total of nine axes of data. Its average recognition accuracy rate reaches 95.46%. on the whole, the greater the number of sensors, the better the recognition effect. When using three sensors together, the performance is best, and the accuracy rate is more than 90%.

5. Conclusions and Future Work

With the development of wearable inertial sensing devices, pattern recognition based on inertial sensors has been studied in various fields and it has gradually become a hot research direction. Some researchers have also used inertial sensors in combination with other types of sensors and have achieved good results in their experiments [26-28]. In this paper, we use the inertial sensing data of smart watch to recognize some important table tennis actions and we did a series of comparative experiments in the case of different sensor combinations and different forms of data organization. Finally, we tested a single experimental individual data and achieved good recognition results. We analyzed all the experimental results and reached the corresponding conclusions.

In addition, we also have some deficiencies in the experimental design and experimental process: a) The individual's actions are not standard enough, and there are obvious differences in individual action features; b) The collected acceleration data contains gravity, which may affect the recognition performance; c) The amount of data is not enough. We collected a total of 12 subjects and could not effectively test the deep differences in action features between individuals.

In future work, we will study action recognition and human behavior feature mining based on inertial sensors more deeply. On the one hand, we will improve the accuracy of data, increase the amount of data collection, and use relevant theories to make more reasonable planning of data sets; On the other hand, we will optimize the recognition model and make exploration and comparison of relevant theories in-depth.

This work is supported by the Major transverse project and the Recruitment Program of Global Experts, under grant SWU41015718 and SWU20710953.

References

[1]. Brand M, Oliver N, Pentland A. Coupled hidden Markov models for complex action recognition[C]// Computer Society Conference on Computer Vision & Pattern Recognition. International Conference on IEEE, 1997: 994-999.

[2]. Poppe R. A Survey on Vision-Based Human Action Recognition. Image and Vision Computing, 2010, 28(6): 976-990.

[3]. Ermes M, Parkka J, Mantyjarvi J, et al. Detection of daily activities and sports with wearable sensors in controlled and uncontrolled conditions[J]. IEEE Transactions on Information Technology in Biomedicine,2008,12 (1): 20-26.

[4]. Shoaib M, Bosch S, Incel O D, et al. A survey of online activity recognition using mobile phones. [J]. Sensors, 2015, 15(1): 2059-2085.

[5]. Wang W F, Yang C Y, Wang D Y. Analysis of Movement Effectiveness in Badminton Strokes with Accelerometers[M]. Genetic and Evolutionary Computing. Yangon, Myanma, 2016: 95-104.

[6]. Dadashi F, Arami A, Crettenand F, et al. A hidden Markov model of the breaststroke swimming temporal phases using wearable inertial measurement units[C]. Body Sensor Networks (BSN). Cambridge, MA, USA, 2013: 1-6.

[7]. Hsu Y L, Wang J S, Lin Y C, et al. A wearable inertial-sensing-based body sensor network for shoulder range of motion assessment[C]. International Conference on Orange Technologies (ICOT). Tainan, Taiwan, 2013: 328-331.

[8]. Caldara M, Comotti D, Galizzi M, et al. A Novel Body Sensor Network for Parkinson's Disease Patients Rehabilitation Assessment[C]. Wearable and Implantable Body Sensor Networks (BSN). Zurich, Switzerland, 2014: 81-86.

[9]. Zhang H, Zhang Z Y. Human Motion Capture System Based on Distributed Wearable Sensing Technology[C]. Wireless Communication and Sensor Network (WCSN). Wuhan, China, 2014: 383-390.

[10]. Wang X, Yang L T, Xie X, et al. A Cloud-Edge Computing Framework for Cyber-Physical-Social Services[J]. IEEE Communications Magazine, 2017, 55(11):80-85.

[11]. Vasilescu M A O. Human motion signatures: analysis, synthesis, recognition[C]// International Conference on Pattern Recognition, 2002. Proceedings. IEEE, 2002:456-460 vol.3.

[12]. Lawrence S, Giles C L, Tsoi A C, et al. Face recognition: A convolutional neural-network approach. IEEE Transactions on Neural Networks, 1997, 8(1): 98-113.

[13]. Neubauer C. Evaluation of convolutional neural networks for visual recognition. IEEE Transactions on Neural Networks, 1998, 9(4): 685-696.

[14]. Lin Min, Chen Qiang, Yan Shui-Cheng. Network in network. arXiv:1312.4400v3,2013.

[15]. Xu Chun-Yan, Lu Can-Yi, Liang Xiao-Dan, et al. Multi-loss regularized deep neural network. IEEE Transactions on Circuits and Systems for Video Technology, 2015, 26(12): 2273-2283.

[16]. C. Randell, H. Muller, Context awareness by analyzing accelerometer data, in: Proceedings of the Fourth International Symposium on Wearable Computers, 2000, pp. 175–176.

[17]. Krizhevsky A, Sutskever I, Hinton G E. ImageNet Classification with Deep Convolutional Neural Networks[J]. Advances in Neural Information Processing Systems, 2012, 25(2): 2012.

[18]. Silver D, Huang A, Maddison C J, et al. Mastering the game of Go with deep neural networks and tree search. [J]. Nature, 2016, 529(7587): 484-489.

[19]. Shao Hong, Chen Shuang, Zhao Jieyi, et al. Face recognition based on subset selection via metric learning on manifold [J]. Frontiers of Information Technology & Electronic Engineering, 2015, 16 (12): 1046 - 1058.

[20]. Yanyan Geng, Ru-Ze Liang, Weizhi Li, et al. Learning convolutional neural network to maximize Pos@ Top performance measure [C]// Proc of European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. 2017.

[21]. Yang J B, Nguyen M N, San P P, et al. Deep convolutional neural networks on multi-channel time series for human activity recognition [C]// Proc of the 24th International Joint Conference on Artificial Intelligence. 2015: 25-31.

[22]. Donahue J, Hendricks L A, Guadarrama S, et al. Long-term recurrent convolutional networks for visual recognition and description [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. 2015:2625-2634.

[23]. Nair V, Hinton G E. Rectified Linear Units Improve Restricted Boltzmann Machines Vinof Nair[C]// International Conference on Machine Learning. 2010: 807-814.

[24]. Gu Jiu-Xiang, Wang Zhen-Hua, Jason Kuen, et al. Recent advances in convolutional neural networks. arXiv: 1512.07108v5, 2017.

[25]. Sainath T N, Mohamed A, Kingsbury B, et al. Deep convolutional neural networks for LVCSR//Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing. Vancouver, Canada, 2013: 8614-8618.

[26]. Neha D, Nasser K. Real-Time Continuous Detection and Recognition of Subject-Specific Smart TV Gestures via Fusion of Depth and Inertial Sensing. IEEE Access, 2018: 7019-7028.

[27]. Qin Z, Lihao N, Qian W. Robust Gait Recognition by Integrating Inertial and RGBD Sensors. IEEE Transactions on Cybernetics, 2018: 1136-1150.

[28]. Enrique Garcia-Ceja, Carlos E. Galván-Tejada, Ramon Brena. Multi-view stacking for activity recognition with sound and accelerometer data. Information Fusion 40 (2018) 45–56.