RECOGNITION OF BADMINTON STROKE ACTIONS USING SENSOR TECHNOLOGY

Yixin LIU[1]

The recognition of badminton stroke actions is conducive to the guidance of badminton technology. This paper briefly introduced the basic process of the sensor-based badminton stroke recognition and then the support vector machine (SVM) algorithm and convolutional neural network (CNN) algorithm used for recognizing strokes. Finally, a test was carried out taking the badminton team members of Chengdu College of University of Electronic Science and Technology of China. The results showed that there were significant differences between drive, lift, and block in the accelerated speed and acceleration variance collected by an acceleration sensor, which could be used in the recognition of stroke actions; the performance of the CNN algorithm was better than that of the traditional SVM algorithm and the random forest algorithm in recognizing acceleration information collected by sensors.

Keywords: sensor; badminton; convolutional neural network; support vector machine.

1. Introduction

Badminton is widely praised as a ball game. Although badminton does not need a simple and small field like table tennis, its antagonism is not inferior to table tennis. As the field of badminton is large and the net in the center is high, the people who participate in badminton need to move their bodies substantially to achieve the effect of physical exercise [1]. Badminton also has competitive events. In order to achieve better results, athletes participating in competitive events need continuous training. In addition to conventional physical training, the training also includes various badminton swing exercises [2]. Traditional swing skill training mainly depends on the guidance of experienced coaches. Coaches observe the athletes' swing action and calibrate their action according to experience. However, the efficiency and effectiveness of this training method depend on the level of the coach, and it is more suitable for one-to-one training [3]. Most of the time, a coach will train more than one student; thus, individual training can not be made for each student. With the development of the intelligent sports industry, coaches can also use motion capture systems to collect the movement data of athletes, such as motion, state, and amount of exercise, and make personalized and

[1] B.P.E., Sports Department, Chengdu College of Electronic Science and Technology of China, Chengdu, Sichuan 611731, China, e-mail: huyi081@163.com

reasonable suggestions for athletes through the analysis of movement data. Liu et al. [4] proposed an algorithm for recognizing temporal among actions and used the recognized patterns to represent actions. Experimental results showed that the method could recognize activities with high accuracy from temporal patterns, and temporal patterns could be effectively used as a middle-level feature of activity representation. Chen [5] proposed a sensor fusion method-based real-time human motion recognition system. The results of offline and real-time experiments verified the effectiveness and real-time throughput of the system. This paper briefly introduced the basic process of the sensor-based badminton stroke recognition and the support vector machine (SVM) algorithm and the CNN algorithm used for recognizing stroke actions. Finally, the badminton team members of Chengdu College of University of Electronic Science and Technology of China were taken as subjects for testing.

2. Motion recognition algorithm based on sensors

2.1. The basic process of stroke recognition using sensors

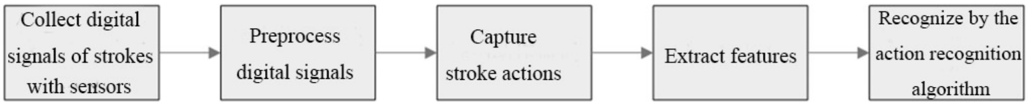


Fig. 1 The basic flow of the sensor-based badminton stroke recognition

① Firstly, the digital signals of badminton players' movements are collected by inertial sensors [6] when they hit the ball. The inertial sensor used is a three-axis acceleration sensor [7], which is used for collecting the digital signals of the three-axis acceleration of badminton players that changes with time when they hit the ball.

② Pre-processing is performed [8], and the formula of filter processing [9] is as follows:

$$\begin{cases} Y(n) = mX(n) + (1-m)Y(n-1) \\ \Delta = Y(n) - Y(n-1) > \Delta_a \\ m = \left(1 - \dfrac{\Delta_a}{\Delta}\right) k_0 \end{cases}, \tag{1}$$

where $Y(n)$ is the output of the current sampling value in the digital signal after filtering, $X(n)$ is the current sampling value of the digital signal, $Y(n-1)$ is the output of the previous sampling value in the digital signal of after filtering, m is the filtering coefficient, Δ is the difference of the output values between the previous and sampling values, Δ_a is the threshold of the motion state, and k_0 is a default parameter [10]. The formula of the composition of accelerations in three directions is:

$$\begin{cases} A_t = \sqrt{a_{xt}^2 + a_{yt}^2 + a_{zt}^2} \\ D_t = (|a_{xt}| - A_t)^2 + (|a_{yt}| - A_t)^2 + (|a_{zt}| - A_t)^2 \end{cases}, \quad (2)$$

where A_t is the resultant acceleration obtained by combining the accelerations in three directions at time t, a_{xt}, a_{yt}, and a_{zt} are the acceleration signal in three directions at time t, and D_t is the variance of the resultant acceleration at time t.

③ After preprocessing the digital signal collected by the sensor, the swing motion is captured. The reason is that the swing action has continuity. Before recognizing the action, the machine needs to divide the action first. In the process of badminton confrontation, the ball is usually swung only once in a round, and then the next swing will be made after the opponent hits back. In this process, the players' actions will pause for a while. The data of the swing action are captured according to the significant difference between the acceleration in the pause period and the swing period [11]. In this paper, a threshold value is set for the acceleration variance of stroke actions to collect the stroke data. The judgment formula is:

$$D_t - D_{t\pm m} \geq U, \quad (3)$$

where $D_{t\pm m}$ is the acceleration variance at m moments before and after D_t and U is the designed threshold. When the difference between the acceleration variance at time t and the acceleration variance at m moments before and after D_t is larger than U, the acceleration at that moment and the acceleration variance are determined as the digital information of stroke actions.

④ The feature of the captured swing is extracted and identified by using a motion recognition algorithm. The detailed recognition algorithm is described in the following.

2.2. Motion recognition algorithm

2.2.1. Support vector machine-based motion recognition algorithm

The recognition algorithm of badminton stroke actions can be regarded as a classification algorithm, which classifies the motions corresponding to the digital signals with the same characteristics into one kind of stroke action. The SVM classification function [12] is:

$$f(x) = \mathrm{sgn}\left(\sum_{i=1}^{n} \alpha_i y_i K(x_i, x) + b\right), \quad (4)$$

where α_i is the Lagrange coefficient, x_i is the feature vector of sample i, y_i is the category label of sample i, $K(x_i, x)$ is a kernel function, which is used for mapping feature vector to the high-dimensional space to reduce the non-linear degree of sample feature distribution, and b is a bias term.

The addition and deletion of non-support vector sample points in the sample data set will not affect the SVM model; thus, the SVM model has good recognition accuracy for small sample classification. SVM is more suitable for binary classification in principle, and the main purpose of this paper is to recognize stroke actions according to the acceleration data collected by the acceleration sensor. There are more than two kinds of stroke actions; therefore, it is necessary to combine different SVMs that recognize different stokes to realize the recognition of multiple kinds of stroke actions. In addition, the SVM model will design large-scale matrix calculation when solving the decision function, the amount of calculation increases with the increase of the number of samples, and a small number of wrong boundary vector points in the sample set will seriously affect the accuracy of the SVM model.

2.2.2. Convolutional neural network-based motion recognition algorithm

The convolution layer [13] of the CNN algorithm is the core structure of the whole network, which performs convolution operation through convolution kernels with fixed specifications. The formula of convolution is:

$$x_j^l = f\left(\sum_{j \in M} x_i^{l-1} \cdot W_{ij}^l + b_j^l\right), \tag{5}$$

where x_j^l is the feature map obtained by the convolution of convolution kernels, x_i^{l-1} is the feature output after the last convolution and pooling, W_{ij}^l is the weight parameter, b_j^l is the offset, M is the number of convolution kernels of the l-th convolution layer, and $f(\bullet)$ is an activation function. Although the convolution operation of the convolution layer only extracts the local sparse features of the original data, after operations of convolution layers and convolution kernels, the local features can form the global features, retaining both the local features and the global features, and each convolution operation in the convolution layer uses the same convolution kernel, which reduces the number of parameters required in training.

The function of the pooling layer is to down-sampling the convoluted features of the convolution layer, i.e., compress the data. The common pooling operations include mean pooling and maximum pooling. The advantage of the pooling layer is that it compresses the features extracted from the convolution

layer, reducing the amount of calculation and avoiding over-fitting; however, the pooling operation will not change the basic features.
The basic training process of the CNN-based stroke recognition model is as follows.

① First, according to the flow steps in Fig. 1, the acceleration signal data of stokes are collected by sensors and preprocessed to capture swing action.

② The processed data are input into CNN for layer-by-layer calculation until the calculation result is obtained in the output layer [14].

③ The recognition result is determined by an objective function. If the recognition result was consistent with the setting, the training ends. If the recognition result is not consistent with the setting, the parameters in the convolution layer are adjusted reversely. The objective function used in this paper is a cross-entropy function [15]:

$$H = -\sum_i y_i' \log(y_i) \quad , \tag{6}$$

where H is the cross-entropy of the recognition result of CNN and the actual result in the training process, y_i is the recognition result in the process of training, and y_i' is the actual result.

3. Case analysis

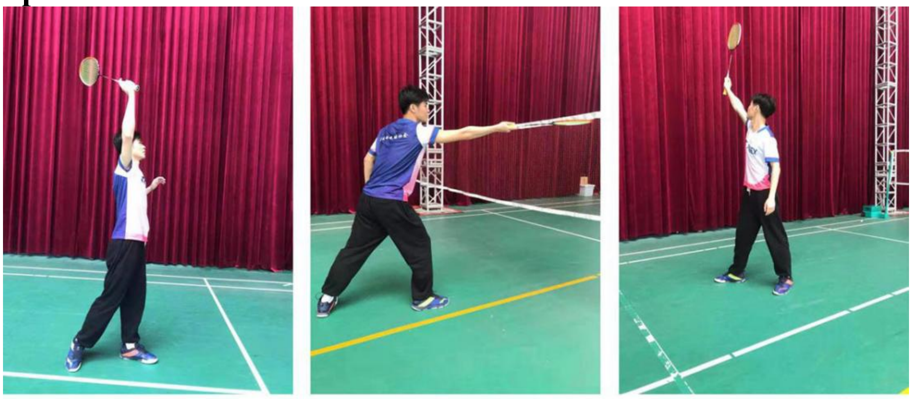
3.1. Subjects

This study took the badminton team members of Chengdu College of University of Electronic Science and Technology of China as the subjects. Ten male athletes with an average age of 20 ± 1 years and an average height of 175 ± 2 cm were selected from the school team. All ten school team members had more than three years of badminton training, had no history of injury, and had no physical injury in the past three months.

3.2. Experimental equipment

In this study, the three-axis acceleration sensor produced by Germany MMF was used as the badminton stroke acceleration signal collection equipment. The three-axis acceleration sensor (KS963B10) had a weight of 8.5 g and a thickness of 8.6 mm. Whether it was installed on the handle of the badminton racket or the wrist, it would not have a significant impact on stroke actions. The frequency of acceleration signal acquisition was 200 Hz.

3.3. Experimental methods



Drawing Picking Blocking

Fig. 2 Three kinds of badminton strokes used in the experiment

KS963B10 three-axis acceleration sensor was used, with a weight of 8.5 g and a thickness of 8.6 mm. It was installed at the bottom of the handle of the badminton racket, as shown in Figure 3. The reason for installing the senor at the end of the handle is as follows. The sensor had a wireless transmission function so that the movement would not be affected by transmission lines. The bottom of the handle was held by a hand, so it would not interfere with movement as much as if it were installed on the arm.

Fig. 3 The installation site of the three-axis acceleration sensor on the racket

The acceleration signal was collected by the sensor at a frequency of 200 Hz. In the sensor module used for the swing acceleration collection, data communication and connection were realized by the inter-integrated circuit (IIC) protocol and master control unit based on the above-mentioned three-axis acceleration sensor. The master control unit transmitted the data to the mobile terminal for storage through the Bluetooth transmitter module of universal asynchronous receiver/transmitter (UART) port.

Ten players made three kinds of stroke actions, including drawing, lifting, and blocking, respectively. In order to avoid the impact of the empty swing on stroke actions, the players returned the badminton ball served by a badminton automatic server with three kinds of stokes. The same stroke action was repeated ten times, as one group; each player did ten groups for each kind of stroke action. The interval between each group was 10 minutes. In total, each player did each kind of stroke action 100 times, and the sample data set contained 1000 strokes, 1000 lifting, and 1000 blocking. In each kind of stroke, 70% of the sample data were selected as the training samples, and the remaining 30% were selected as the testing samples. The specific selection way is as follows. Taking the drawing action as an example, the athlete repeated the drawing action 1000 times; 700 out of the 1000 actions was selected as the training samples, and the remaining 300 actions were selected as the testing samples. Every kind of action was processed in the same way. Finally, there were three kinds of actions in the training samples, 700 for every kind of action, totally 2100; there were also three kinds of actions in the testing sets, 300 for every kind of action, totally 900.

In order to verify the effectiveness of the CNN algorithm in recognizing stroke actions, it was compared with the SVM algorithm and the random forest algorithm.

3.4. Evaluation criterion

Table 1

Confusion matrix

	Determined as positive by the algorithm	Determined as positive by the algorithm
Positive actually	TP	FN
Negative actually	FP	TN

It was difficult to comprehensively evaluate the effectiveness of the stroke recognition algorithm only by using the accuracy; therefore, the recognition performance of the algorithm was evaluated by the confusion matrix. The confusion matrix is shown in Table 1, then the corresponding performance index calculation formula is:

$$\begin{cases} P = \dfrac{TP}{TP + FP} \\ R = \dfrac{TP}{TP + FN} \\ F = \dfrac{2TP}{2TP + FP + FN} \end{cases}, \tag{7}$$

where P refers to the accuracy, R refers to the recall rate, F is a comprehensive index of accuracy and recall rate, TP refers to the number of positive samples which were determined as positive, FP refers to the number of negative samples which were determined as positive, FN refers to the number of positive samples which were determined as negative, and TN refers to the number of negative samples which were determined as negative.

3.5. Experimental results

The stroke, lifting, and blocking of one athlete was taken as an example, and the resultant acceleration and acceleration variance of his arm are shown in Figs. 4 and 5. It was seen from Figs. 4 and 5 that the resultant acceleration and acceleration variance of the three strokes had relatively similar tendencies with the change of time in one cycle of stroke, i.e., keeping stable firstly, then suddenly increasing, and finally rapidly decreasing. Although the variation tendencies of acceleration and acceleration variance of the three strokes were similar, there were significant differences in the specific numerical changes. The acceleration and acceleration variance of lifting in the middle of an action cycle increased the least, while the increase of drawing was the largest.

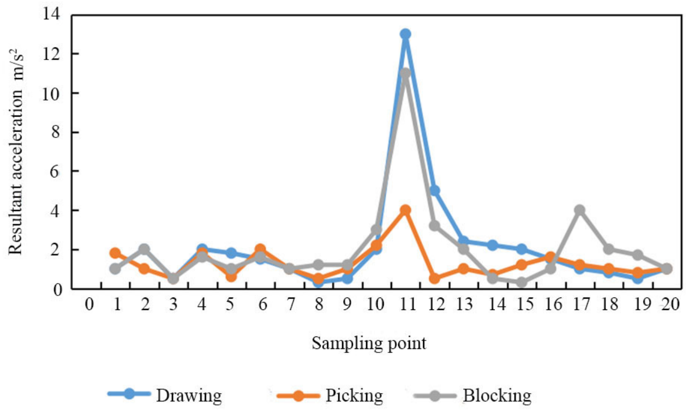


Fig. 4 The resultant acceleration of three strokes in one cycle

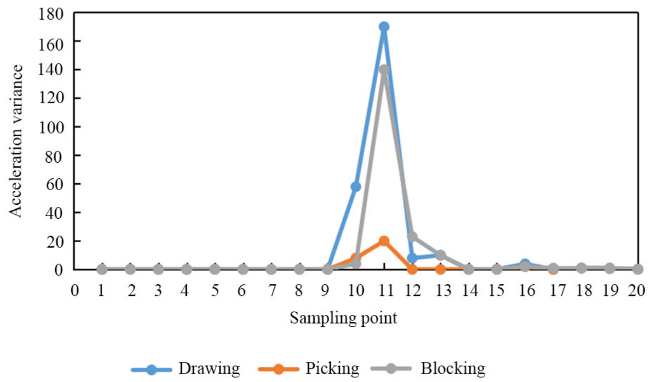


Fig. 5 The angular velocity variance of three strokes in one cycle

It was seen from Fig. 6 that the recognition performance of the random forest algorithm was the worst, and that of the CNN algorithm was the best. The reason was that the random forest algorithm and the SVM algorithm needed to extract the features of the stroke data manually when classifying sensor data, and these features might not be suitable for the random forest algorithm and the SVM algorithm. When the CNN algorithm classified sensor data, it did not need to extract features in advance but directly classified the preprocessed original data. In the process of classification, the features of the original data have been extracted in the convolution layer. These features include not only local features but also global features. The existence of the pooling layer not only reduced the amount of calculation but also filters the extracted features to a certain extent. Thus, the CNN algorithm had the highest recognition performance for strokes.

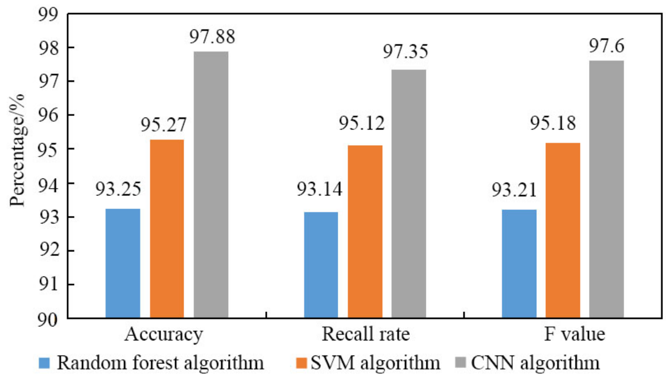


Fig. 6 The recognition performance of three recognition algorithms for strokes

4. Conclusion

This paper briefly introduced the basic process of the sensor-based badminton stroke recognition and SVM algorithm and CNN algorithms used for recognizing stroke actions. Finally, the badminton team members of Chengdu

College of University of Electronic Science and Technology of China were tested. The results are as follows. The variation trend of acceleration and acceleration variance with time in one action cycle of the three strokes was similar, i.e., keeping stable first, then suddenly change, and finally becoming stable. In addition, the sudden variation degree of drawing was the largest, blocking was the second, and picking was the smallest. The CNN algorithm had the highest recognition accuracy, recall rate, and F-value, followed by the SVM algorithm, and the random forest algorithm had the worst recognition performance.

REFERENCES

[1] *G. M. Zhu, L. Zhang, P. Y. Shen and J. Song*, "An Online Continuous Human Action Recognition Algorithm Based on the Kinect Sensor", Sensors, **vol. 16**, no. 2, 2016, pp. 161.

[2] *J. Hong, B. Cho, Y. W. Hong and HByun*, "Contextual Action Cues from Camera Sensor for Multi-Stream Action Recognition", Sensors, **vol. 19**, no. 6, 2019, pp. 1382-.

[3] *Y. Gu, M. Liu, W. Sheng, Y. Ou and Y. Li*, "Sensor fusion based manipulative action recognition", Auton. Robot., no. 3, 2020, pp. 1-13.

[4] *Y. Liu, L. Nie, L. Liu and D. S. Rosenblum*, "From action to activity: Sensor-based activity recognition", Neurocomputing, 2016, pp. 108-115.

[5] *C. Chen, R. Jafari and N. Kehtarnavaz*, "A Real-Time Human Action Recognition System Using Depth and Inertial Sensor Fusion", IEEE Sens. J., **vol. 16**, no. 3, 2016, pp. 773-781.

[6] *G. Zhu, Z. Liang, P. Shen and J. Song*, "An Online Continuous Human Action Recognition Algorithm Based on the Kinect Sensor", Sensors, **vol. 16**, no. 2, 2016, pp. 161.

[7] *C. Rupprecht, C. Lea, F. Tombari, N. Navab and G. D. Hager*, "Sensor substitution for video-based action recognition", IEEE/RSJ International Conference on Intelligent Robots & Systems, 2016, pp. 5230-5237.

[8] *H. Mohammadzade, S. Hosseini, M. R. Rezaei-Dastjerdehei and M. Tabejamaat*, "Dynamic Time Warping-Based Features With Class-Specific Joint Importance Maps for Action Recognition Using Kinect Depth Sensor", IEEE Sens. J., **vol. PP**, no. 99, 2021, pp. 1-1.

[9] *Z. Ahmad and N. Khan*, "Inertial Sensor Data To Image Encoding For Human Action Recognition", IEEE Sens. J., **vol. PP**, no. 99, 2021, pp. 1-1.

[10] *Y. Guo, D. Tao, W. Liu and J. Cheng*, "Multiview Cauchy Estimator Feature Embedding for Depth and Inertial Sensor-Based Human Action Recognition", IEEE T. Syst. Man Cy. S., **vol. PP**, no. 99, 2016, pp. 1-11.

[11] *G. Xiao*, "Research on athlete's action recognition based on acceleration sensor and deep learning", J. Intell. Fuzzy Syst., **vol. 40**, no. 2, 2021, pp. 2229-2240.

[12] *M. Naveenkumar and A. Vadivel*, "3-D Projected PCA Based DMM Feature Fusing with SMO-SVM for Human Action Recognition - ScienceDirect", Proc. Comput. Sci., **vol. 89**, 2016, pp. 759-763.

[13] *A. Ullah, K. Muhammad, I. U. Haq and S. W. Baik*, "Action recognition using optimized deep autoencoder and CNN for surveillance data streams of non-stationary environments", Future Gener. Comp. Sy., **vol. 96**, no. JUL., 2019, pp. 386-397.

[14] *M. Lu, Y. Hu and X. Lu*, "Driver action recognition using deformable and dilated faster R-CNN with optimized region proposals", Appl. Intell., **vol. 50**, no. 3, 2020.

[15] *X. Ouyang, S. Xu, C. Zhang, P. Zhou, Y. Yang, G. Liu and X. Li*, "A 3D-CNN and LSTM based Multi-task Learning Architecture for Action Recognition", IEEE Access, **vol. 7**, no. 99, 2019, pp. 40757-40770.