Research on the Application of Artificial Intelligence Technology in the Field of Sports Refereeing

Xin Chen
Sports Department, Shenyang Jianzhu University, Shenyang 110168, China
1073907651@sjzu.edu.cn

Abstract. This article uses artificial intelligence machine vision technology to design a method for judging violations of passing passes. First, use Adaptive Enhanced Variant Algorithm (GAB) to train the classification model to complete the detection of sports (badminton and racket); combine the detection area to extract the sphere center coordinates are used as the target point, and the detected racket part is constructed as the target area. Then the position relationship between the two is defined, and the judgment key frame is extracted according to the relative position change relationship between the target point and the target area; finally, the image angle representation method is used to establish the state judgment model takes the state of the racket in the judgment key frame as an input, and quantizes the output result as an angle value, so as to realize the judgment on whether to serve the ball.

Keywords: Artificial intelligence, sports, badminton, refereeing.

1. Introduction
The current badminton referee system is a completely subjective referee system, and the result of each arbitration is completely dependent on the eyes of the referee or observer to determine the specific location of the badminton. Therefore, there are often controversies during badminton games. Even if the video playback method is used, the human eye's observation of the image is ultimately used as the judge's conclusion. It is difficult to accurately determine the position of the badminton and cannot give an objective conclusion. According to the above analysis, the application of artificial intelligence technology to the field of competitive sports is an inevitable trend in the future. This article mainly focuses on the game serve scene [1]. Considering that the prescribed action in the scene usually contains multiple body parts, it is necessary to recognize and judge multiple parts to realize the illegal action detection. The complexity is high, and the simple action detection may reach If the application scenario does not meet the requirements for the detection of illegal actions, another approach is needed.
2. Target detection technology

2.1. Motion capture technology
Motion capture technology refers to the use of sensors, motion capture equipment, data transmission equipment, and data processing equipment to map the motion trajectory of an object to a computer, and convert it into a corresponding motion trajectory model according to a specific algorithm. Generally, a complete motion capture system should include several aspects as shown in Figure 1.

![Figure 1. Motion capture technology](image)

The sensor can realize the perception of the movement state of the object, including position information such as the speed, acceleration, angle and height of the object movement. The signal capture device is mainly used to establish a connection with the sensor and capture the information of the sensor [2]. The data transmission device is the bridge between the signal capture device and the computer, and can transmit the information of the signal capture device to the computer. Data processing equipment is mainly used to process the movement information of objects, and analyse and reproduce the collected movement information in a computer system.

2.2. System development platform
The posture motion capture is a variety of sensors including infrared sensors, high-definition cameras, microphone arrays, and digital signal processors. The device can sense the original data stream, track the skeleton image of one or two users within its field of view, and establish a corresponding somatosensory operation program, which satisfies the motion capture technology in the badminton training process requirements.

2.3. Digital image feature technology
For any pixel $p(x, y)$ in a certain area of an image, assuming its gray level is $g$, then remove the center point in a window of $w$, and the gray level of the remaining points can be described as: $g_{w}, ..., g_{w-1}$, then the center pixel and the adjacent pixels The relative relationship of can be described as
Taking the gray level of the central pixel as the threshold, and binarizing \( b(g_z - g_o), \ldots, b(g_z - g_{w-1}) \), which is defined as follows:
\[
b(x) = \begin{cases} 
1 & x < 0 \\
0 & x \geq 0 
\end{cases}
\]

Through this method, a binary number is obtained, and the corresponding characteristic value is obtained after weighting, which is expressed as a mathematical formula as follows:
\[
LBP_{(x,y)} = \sum_{i=1}^{w} b(g_z - g_i) \times 2^{i-1}
\]

Count the LBP values of all pixels, and use the LBP histogram as the representation of the image. Assuming an image with pixel \( A \times B \), the statistical method of its histogram is:
\[
H(k) = \sum_{x=1}^{A} \sum_{y=1}^{B} \left[ LBP_{(x,y)} - k \right]
\]

Among them, \( k \) needs to satisfy \( 0 \leq k < 2^{w-1} \), which represents the vector dimension of the LBP histogram of the entire image. The original LBP feature calculation window is defined as a \( 3 \times 3 \) rectangular window, and the calculation process is shown in Figure 2 below:

![Figure 2. Original LBP feature extraction method](image)

With the increase of sampling pixels, a large amount of data not only affects the calculation speed, but also causes the final histogram distribution to be too sparse, which is not conducive to describing the entire image. In order to reduce the dimensionality of large-scale LBP features, uniform LBP was born. This method divides the LBP features of the entire image into two categories: uniform mode and non-uniform mode [3]. The LBP features classified as uniform mode have the following definitions, whose value \( U_{LBP} \) is less than or equal to 2:
\[
U_{LBP} = \sum_{i=1}^{w-1} \left| b(g_z - g_i) - b(g_z - g_{\text{med}(i+1)}) \right|
\]

The value \( U_{LBP} \) can be understood as the number of transitions from 0 to 1 and 1 to 0 in the binary number corresponding to the LBP feature of a certain circular domain does not exceed two times. Once it exceeds two times, it is classified as a non-uniform mode. This method the dimensionality reduction of LBP features is realized. Originally, the total number of LBP features in the circular domain with the sampling point \( w \) is \( 2^w \). Using uniform LBP can reduce the total to \( w \times (w - 1) + 2 \), and then including the non-uniform mode, the total number of features is \( w \times (w - 1) + 3 \). This article will extract the LBP features of badminton and racket to train the classification model separately, and test the detection effect in different environments.
3. Realization of badminton target detection

3.1. Target detection data processing

3.1.1 Data set establishment. The training sample is divided into two parts: positive sample and negative sample. The positive sample is the object that needs to be detected. In the badminton classification model, the positive sample is the picture that only contains the shuttlecock, and the negative sample can theoretically be anything that does not contain the shuttlecock. Pictures, but this approach is not only useless, but reduces the relevance and increases the time spent in training, with half the effort. The correct approach should be to select the corresponding negative samples for the application scenarios. In this topic, the negative samples can be selected indoor, outdoor badminton courts or badminton court-like venues, venues and venues in different time periods or under different lighting conditions [4]. All around are excellent negative sample materials. In order to select a suitable image data set, I compared multiple large image data set sites: Image eNET, SUN, Object Categories, etc., and found that these sites did not fully support badminton application scenarios, and there was no ready-made data set available. This requires building a badminton related data set by yourself.

For the racket classification model, the positive sample is the picture that only contains the racket, and the negative sample is composed of non-racquet pictures. Taking into account the correlation with the application scenario, the negative sample is still composed of indoor and outdoor badminton courts, badminton courts and related pictures. A total of 20655 sheets. The negative sample of the racket classification model is slightly different from the negative sample of the badminton classification model. The negative sample of the former can contain pictures of badminton, and the negative sample of the latter can contain pictures of rackets. The two are not completely common. In order to ensure that the trained classification model achieves the optimal classification effect, the ratio of positive and negative samples should be maintained at 1:4, and a total of 5129 racket pictures taken in different directions should be taken.

3.1.2 Sample retreatment. Before training the classification model, the positive sample pictures need to be further processed to generate the positive samples for training. The first is to unify the specifications of the positive sample pictures. Feature extraction for pictures with too large size will result in too many features and affect the training time. In this topic, the positive samples are uniformly set to the pixel size, and the negative samples do not need to be unified in size. Secondly, convert all positive and negative samples into grayscale images [5]. The purpose of this is to remove the colour information in the image and reduce training time. The purpose of these two treatments is to improve training efficiency. The positive samples of badminton and rackets with uniform size are shown in Figure 3 below:

![Figure 3. An excerpt of a positive sample of badminton after uniform processing](image-url)
3.2. Cascade model training and detection

After the positive and negative sample data passes through the classification model, two different results will be produced respectively. The combination of these four different results is a measure of the performance of the classification model. For positive sample data, there are two results: the first, the positive sample is correctly classified by the classification model, called TP; the second, the positive sample is classified by the classification model as background (i.e., negative sample), which is equivalent to being the missing points are called FN. For negative sample data, there are two results: the first type, the negative sample is correctly classified by the classification model, called TN; the second type, the negative sample is classified as the target (i.e., the positive sample) by the classification model, which is equivalent to the error Points, called FP. The accuracy and recall rate of two common performance measurement indicators of classification models are defined as follows:

\[
\text{Precision} = \frac{TP}{TP + FP} \quad (6)
\]

\[
\text{Recall} = \frac{TP}{TP + FN} \quad (7)
\]

Among them, accuracy is a measure of how many classified results are correctly classified. The recall rate indicates how many positive samples are classified among all positive samples, and it measures the ability to classify positive samples. In training using AdaBoost, more attention is paid to measuring the positive samples and negative samples separately [6]. The combination of the two results of the positive sample classification is defined as the hit rate, and the combination of the two results of the negative sample classification is defined as the false alarm rate or the false detection rate:

\[
\text{HitRate} = \frac{TP}{TP + FN} \quad (8)
\]

\[
\text{FalseAlarm} = \frac{FP}{FP + TN} \quad (9)
\]

It can be seen that Hit Rate is equal to Recall, which measures the ability of the classification model to pass positive samples, and False Alarm measures the ability of the classification model to pass negative samples. Obviously, the higher the Hit Rate, the better, and the lower the False Alarm, the better.

3.3. Implementation of strong classifier training

The strong classifier trained by AdaBoost is divided by the number of layers (Stage). Each layer represents an AdaBoost strong classifier, and each strong classifier is formed by a linear combination of several weak classifiers. A single AdaBoost weak classifier is a one-layer decision tree structure. The structure diagram of the strong classifier is shown in Figure 4 below:

![Figure 4. Structure diagram of strong classifier](image-url)
4. Experiment and result analysis
On the basis of the first two single probe experiments, preliminary experiments are carried out on the array detection system. The experiment consists of nine fibre couplers to form a detection array. The Gaussian beam emitted by the laser is aligned and coupled with the fibre bundle through laser beam expansion to realize the coupling of the light source to the fibre. The array fibres used are numbered (from top to bottom, from left to right) from 1 to 9. When the detection port fibre is unobstructed, the image obtained by the CCD camera and its histogram. You can see that there are nine points of different brightness in the image. When the detection port fibre is blocked, the image will change accordingly. The inspection experiments on each optical fibre have successfully verified the feasibility of the inspection system. Due to the difference in the coupling processing of the fibre coupler itself and the uneven coupling effect of each fibre due to the limitation of experimental conditions, there are points of different brightness in the original image without obstruction [7]. This experiment mainly compares the changes of the images before and after, so the original image has points with different brightness has no effect on the verification of the experiment. The part of the detection results including the badminton image frames detected during the total video time is shown in Figure 5 below:

Figure 5. Badminton test results

The badminton classification model was tested in three environments. Table 1 shows the statistical results of the number of detected frames, and Table 2 shows the recall statistics when the intersection ratio is greater than 0.5:

**Table 1.** Badminton classification model detection statistics

| Surroundings | Total number of frames | Number of detection frames | Recall rate |
|--------------|------------------------|----------------------------|-------------|
| Natural light | 2700                   | 2472                       | 91.5%       |
| Artificial light | 2700            | 2486                       | 92.1%       |
| Low light    | 2700                   | 2449                       | 90.7%       |

**Table 2.** Badminton test recall rate

| Surroundings | IoU | Number of detection frames | Recall rate |
|--------------|-----|----------------------------|-------------|
| Natural light | 0.5 | 2451                       | 90.8%       |
| Artificial light | 0.5    | 2467                       | 91.4%       |
| Low light    | 0.5 | 2428                       | 89.9%       |

The average recall rate of the badminton classification model trained in this subject is 91.4% under different lighting environments, and the average recall rate is 90.7% when the intersection ratio is greater than 0.5, which has certain performance.
5. Conclusion

Aiming at the problems existing in the current badminton training process, the article studies an auxiliary training system based on posture motion capture technology. The system can use the attitude depth sensor to capture the motion state data, and at the same time use the DWT algorithm to analyse the motion to establish a motion database. In addition, establish corresponding training plans according to the characteristics of different athletes. The content of this research is of great significance for improving the level and efficiency of badminton ball training.

References

[1] Cui, J., Liu, Z., & Xu, L. Modelling and simulation for table tennis referee regulation based on finite state machine. Journal of sports sciences, 35(19) (2017) 1888-1896.
[2] Zhan, Y., & Tan, K. H. An analytic infrastructure for harvesting big data to enhance supply chain performance. European Journal of Operational Research, 281(3) (2020) 559-574.
[3] Rahmad, N. A., Sufri, N. A. J., Muzamil, N. H., & As’ari, M. A. Badminton player detection using faster region convolutional neural network. Indonesian Journal of Electrical Engineering and Computer Science, 14(3) (2019) 1330-1335.
[4] Taborri, J., Palermo, E., & Rossi, S. Automatic detection of faults in race walking: A comparative analysis of machine-learning algorithms fed with inertial sensor data. Sensors, 19(6) (2019) 1461-1469.
[5] Shih, H. C. A survey of content-aware video analysis for sports. IEEE Transactions on Circuits and Systems for Video Technology, 28(5) (2017) 1212-1231.
[6] Rezaii, N., Walker, E., & Wolff, P. A machine learning approach to predicting psychosis using semantic density and latent content analysis. NPJ schizophrenia, 5(1) (2019) 1-12.
[7] Howe, M. S. The world is its own best model: modelling and future pandemic planning in dentistry. British dental journal, 229(11) (2020) 716-720.