2021 13th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA) | 978-1-6654-3892-6/20/$31.00 ©2021 IEEE | DOI: 10.1109/ICMTMA52658.2021.00108

Badminton Video Analysis Based on Player Tracking and Pose Trajectory Estimation

Cai Cuiping
Polus International College, Chengdu 610103, China
E-mail:962671625@qq.com

***Abstract*—The traditional attitude estimation method is still not ideal for occlusion, which often ignores the continuity of human posture itself and the spatiotemporal continuity of human target trajectory. Therefore, this paper considers using the improved CAMSHIFT algorithm to track the athletes, taking badminton as an example, proposes a human posture estimation algorithm based on video. The scheme uses H, S, V to establish the athletes' color model. The traditional histogram is replaced by weighted histogram and instance histogram. The algorithm also estimates the pose of the human body in static image, and establishes the human body tracking based on the measurement of the pose distance between frames. It generates better candidate keys for the key points with lower confidence in each frame, and generates the optimal global pose for each frame by recombining these body parts. Finally, the real-time semantic analysis scheme of badminton video is formed by using the above methods, and the performance indexes of the whole system are comprehensively analyzed through experiments. It ensures the tracking processing speed and reduces the error caused by similar tracking targets, which proves the feasibility and effectiveness**

Keywords-pose trajectory; Camshift; Openpose; badminton athlete; deep learning

I. INTRODUCTION

In recent years, sports games viewing has become an important part of modern people's life, and the importance of sports video is gradually highlighted. The intelligent system based on information technology has played an important role in modern competitive sports, but there is no relevant research on the video semantic analysis technology of the whole badminton game. This paper mainly focuses on the semantic analysis of the two major sports objects in the badminton game - players and badminton. The output semantic information includes the type of action, distance, speed and landing area of badminton players. A player tracking algorithm based on role partition is designed, which continuously locks the target player to be tracked and outputs the player area. After that, from coarse to fine, we design an action type discrimination algorithm based on the key point displacement of human body and an algorithm to detect the movement distance and speed of athletes. Through the two-dimensional attitude estimation of players in the player area, we analyze their body skeleton and key point information to complete the analysis of players' movement types. At the same time, according to the displacement information of the key points and the running time of the system, the distance and speed of the athletes are obtained. Finally, this paper implements a semantic analysis prototype system of real-time video of badminton match, and tests and analyzes the performance indicators of the system. The results show that the system has good performance in real-time, accuracy and stability.

II. RELATED KNOWLEDGE

A. Target Detection Technology Based on Deep Learning

In recent years, with the rapid development of deep learning technology, the target detection algorithm has changed from the traditional algorithm based on manual features to the detection technology based on deep neural network. These algorithms have excellent detection effect and performance on the open target detection data set. ZFNet is illustrated as as an example, which contains five convolution layers. Figure 1 depicts the convolution neural network model ZFNet's feature mapping of each convolution layer visualized by deconvolution network. The feature map of the first and second volume integrations shows that the edge, color and corner features are obvious, and the color, line and edge can be seen clearly in the second layer of the figure. The convolution feature map of the third volume of the integrations shows that the similar texture and translation invariance are obvious, and the grid features in the upper left corner of the figure are special The features can be visualized in the feature map, and the fourth and fifth volume of accumulation begin to show the differences between classes. Then a certain depth is necessary for neural network.

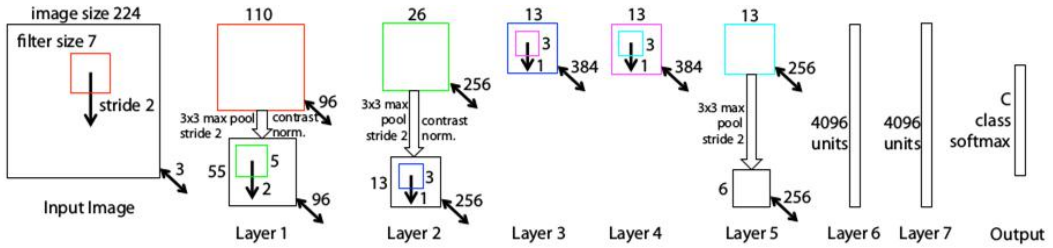


Figure 1. Convolution feature mapping of ZFnet

978-1-6654-3892-6/21/$31.00 ©2021 IEEE
DOI 10.1109/ICMTMA52658.2021.00108

B. 2D human posture recognition

From the point of view of computer vision, the best way of gesture recognition is to find out its movement characteristics. These features can include many aspects, such as the human body's step frequency, stride, facial features, gestures, body posture, walking track, shaking degree and so on. The effect of only relying on computer vision algorithm is often not good, usually combined with some hardware devices. In addition to the combination with hardware, it is a good method to recognize human motion and estimate human posture by obtaining the information of human skeleton and key points. Openpose human posture recognition project is an open source library developed by Carnegie Mellon University based on convolutional neural network and supervised learning and Caffe framework. It can realize the estimation of human motion, facial expression, finger movement and so on. It is suitable for single person and multi person with excellent robustness. It is the first real-time multi person two-dimensional attitude estimation application based on deep learning in the world. The implementation principle is as follows: input an image, extract features through convolution network, get a group of feature maps, and then divide them into two branches, and extract part conformance maps and part affinity fields by CNN network respectively; after getting these two information, we use bipartite matching in graph theory to find part Association is to connect the joint points of the same person. Due to the Vectoriality of PAF itself, the generated even matching is very correct, and finally it is merged into the overall skeleton of a person. Finally, based on PAFS, multi person parsing is calculated, and the multi person parsing problem is transformed into graphs problem, which is called Hungarian problem Hungarian algorithm is the most common algorithm for partial graph matching. The core of this algorithm is to find the augmented path, which is an algorithm to find the maximum matching of bipartite graph with the augmented path.

III. SEMANTIC ANALYSIS OF REAL-TIME VIDEO FOR BADMINTON MATCH

A. Maind Concept of Algorithm

To analyze athletes' sports information, we need a general strategy and step design. After a large number of data analysis and experimental proof, we adopted the research process as shown in figure 2. Firstly, the object of the input HD badminton game video is detected, and the objects that need semantic analysis, i.e. human and badminton, are detected. Then, the role of the personnel is identified, so as to lock the target player and track it continuously. Then, the key points of the target player are detected, so that the position information of each key point is obtained, according to the key points. At the same time, the movement data is counted. In this way, we acquire the output semantics.

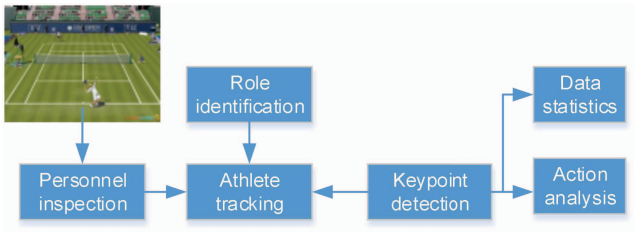


Figure 2. Real time semantic parsing algorithm flow of motion video

B. Badminton Player Tracking Approach

The purpose of badminton player tracking is to obtain the player's location information, and at the same time, combining with the previously obtained field line information to achieve the marking and detection of high-level events. In this section, the improved CAMSHIFT algorithm is used to track badminton players. Another disadvantage of the traditional CAMSHIFT algorithm is that it is a semi-automatic tracking algorithm, which needs to manually select the initial position and size of the search window. In order to use CAMSHIFT algorithm to track athletes automatically and robustly in sports video, this paper improves CAMSHIFT in the following aspects.

(1) By combining the athlete detection process with CAMSHIFT algorithm, the detected athlete area is regarded as the initial search window of CAMSHIFT algorithm, so as to avoid the process of manually selecting the initial search window and realize the athlete's automatic tracking.

(2) The information of H, S and V in HSV space is used to build the color model for athletes. H, S, V are used to generate multi-dimensional color histogram, and the back projection of multi-dimensional color histogram is used to transform the video image into probability distribution map. Because the probability distribution map contains rich feature information, it can be used to locate and track athletes more accurately. Moreover, since the inverse projection of multi-dimensional histogram is linear to the number of feature spaces, the increase of computational complexity caused by increasing the dimension of feature space is relatively small.

(3) The method of weighted histogram and scale histogram is used to improve CAMSHIFT algorithm. The sample points around the target are often affected by noise, which reduces the reliability. In the athlete's area, the farther away from the center of the area, the lower the reliability of the pixel, the more likely the pixel is to be covered by other objects or the pixel belongs to the background. Therefore, it is reasonable to give different weights to the pixels of different positions in the region when calculating the histogram of athletes' region. The closer the distance to the center of the region is, the larger the weight of the pixel is. The calculation method of the weighted histogram of the athlete area is as follows: the multi-dimensional color histogram is generated by H, S and V, and the video image is transformed into probability distribution map by the back projection of the multi-dimensional color histogram. Since the probability distribution map contains rich feature information, it can be used to locate and track athletes more

accurately. Moreover, since the inverse projection of multi-dimensional histogram is linear to the number of feature spaces, the increase of computational complexity caused by increasing the dimension of feature space is relatively small.

$$q_u = \sum_{i=1}^{n} k(\| x_i^* \|^2)\delta[c(x_i^*) - u]$$

where x_i^* is the pixel of different location, $k(x)$ is monotonically decreasing kernel function whose function is assigning a little weight to the pixel far from the window center. The simplest kernel function for generating weighted histograms is depicted as

$$k(r) = \begin{cases} 1-r, & r \le 1 \\ 0, & else \end{cases} \quad (1)$$

The elliptical frame is used to represent the tracking target. In most cases, the frame contains background pixels or neighboring objects' pixels, so the histogram generated by the color information in the elliptical frame also contains interference features. In this case, using the weighted histogram to transform the video image into the color probability distribution map cannot locate the athletes well. In this paper, the scale histogram is used to solve the problem of interference between background and adjacent objects. In the process of scale histogram generation, the background features and the features of adjacent objects are given lower weight and the target objects are given higher weight. The process of calculating the scale histogram is as follows:

Firstly, the color histogram is calculated for the background area of the outer neighborhood of the athlete area as

$$B_u = \sum_{i=1}^{n} k(x)(\| x_i^* \|^2)\delta[c(x_i^*) - u] \quad (2)$$

The kernel function is

$$k(r) = \begin{cases} ar, & 1 < r \le h \\ 0, & else \end{cases} \quad (3)$$

where a is scale factor, and h is histogram area of calculating the window width which is larger than the athlete area. Then the weight parameter of background area is computed as:

$$\{W_u = \min(\frac{B*}{B_u}, 1)\}_{u=1,2,...,m} \quad (4)$$

where $B*$ is the minimum color index in $\{B_u\}_{u=1,2,...,m}$. The scale histogram can be acquired as

$$q_u = w_u \sum_{i=1}^{n} k(\| x_i^* \|^2)\delta[c(x_i^*) - u] \quad (5)$$

(4) For the occlusion problem between athletes in sports video, this paper uses the method of search window contrast to solve it. In sports video, athletes often block each other. When the target is blocked, its corresponding search window will gradually become smaller; when the target is connected with a large area of interference, its corresponding search window will become larger. The way to solve the occlusion problem is to compare the search window of a moving object according to a certain time interval. When the size of the search window changes more than a certain proportion, it indicates that there is an abnormal situation. At this time, the athlete detection algorithm is used to detect the athlete area again, and it is used as the initial search window of the CAMSHIFT algorithm, and then the CAMSHIFT algorithm is used to continue the search athletes.

C. *Analysis of Real-time Motion Semantics of Badminton Players Based on Deep Learning*

At present, the human body attitude estimation method based on deep learning is the most promising method among many human body attitude estimation methods. It is not only in the research field of human body attitude estimation, but also in other fields. At present, due to the rapid development of computer hardware equipment, the calculation cost is greatly reduced, and the research based on convolutional neural network is becoming more and more popular, and many fruitful achievements have been achieved However, the general research based on convolutional neural network is still difficult to meet the requirements in real-time. In this paper, we need to apply the human posture estimation to the golf swing action comparison and analysis system in the future, which still has certain requirements on the calculation time. Therefore, in order to ensure that the algorithm can be completed in a short time, this chapter selects Openpose. As a human bone joint point detector, the algorithm is shown in figure 3. The graph is the whole network architecture of open pose. The network structure is divided into two branches. The upper branch network is used to predict the confidence graph l of bone joint points, and the lower branch network is used to predict the partial affinity vector field s.

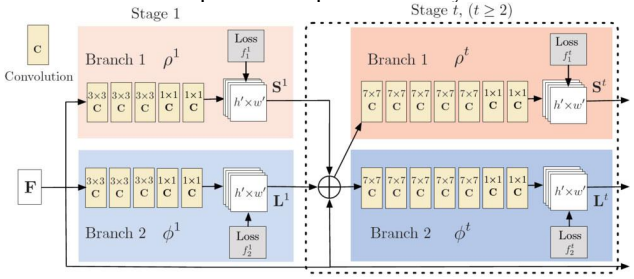


Figure 3. Openpost network architecture

The method of inter frame attitude distance calculation is used to realize the human body attitude tracking in the video. The attitude of the previous frame is associated with the attitude of the next frame. The inter frame attitude distance is the distance between one attitude p_1 in one frame and another attitude p_2 in the next frame. By importing optical flow information to measure the possibility that the posture in two frames is the same person. p_1^i and p_2^i are the i_{th} bone joint point of p_1 and p_2. The border B_1^i and B_2^i are extracted from p_1^i and p_2^i, and the size is decided

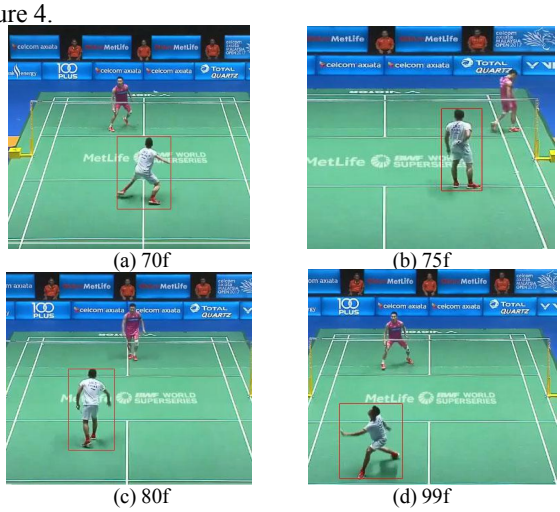
according to standard PCK. Then, Deepatching is used to evaluated the similarity of B_1^i and B_2^i. Due to m_i feature points in B_1^i we can find m_i matching point from B_2^i. Thus, $\frac{n_i}{m_i}$ can be used to represent the similarity of B_1^i and B_2^i. Finally, the distance between the frames are computed as

$$d_c(p_1, p_2) = \sum_i \frac{n_i}{m_i} \quad (6)$$

Based on the measurement of inter frame attitude distance, the correlation matrix of inter frame attitude can be established, and then the tracking information of inter frame human attitude can be established by Kuhn Munkres algorithm.

IV. EXPERIMENTAL ANALYSIS

The badminton video clip tracked in this paper is the selected global game lens including badminton court. The purpose is to obtain the relative position of players, and provide useful player position information for the subsequent use of the badminton court ground line and the relative position of players to obtain high-level semantic annotation. There are a lot of camera movements in the process of badminton video shooting. The background and foreground in the video image are moving, which causes a lot of interference to the players' extraction, especially for the players in the backcourt, the area is very small, often get out of the noise. Therefore, when applying the improved CAMSHIFT algorithm, the semi-automatic method is adopted in this paper, and the position of the initial search box needs to be given manually. If the follow-up improvement can extract two players' positions well, it can realize automatic tracking. This paper selects 70-99 frames of badminton video, and its tracking effect is shown in figure 4.



(a) 70f (b) 75f (c) 80f (d) 99f

Figure 4. Tracking effect in badminton video

The experimental results show that when the improved CAMSHIFT algorithm is used to track the players in the badminton video, a satisfactory tracking effect can be obtained. But when the player moves very fast, sometimes the tracking will fail.We compare the running speed of Openpos network model with that of ZFnet network model. It can be seen that the model can improve the speed obviously while ensuring the accuracy. Through the experiment, it is found that using four key points on the left and right hands of the athlete's arm to detect the swing action has a good effect. The operation result is shown in figure 5. The experimental results show that the algorithm in this paper is better than that in Openpose in every bone joint point, which can verify that the algorithm in this paper is effective, and the algorithm in this chapter can repair the problems of Openpose algorithm to a certain extent.

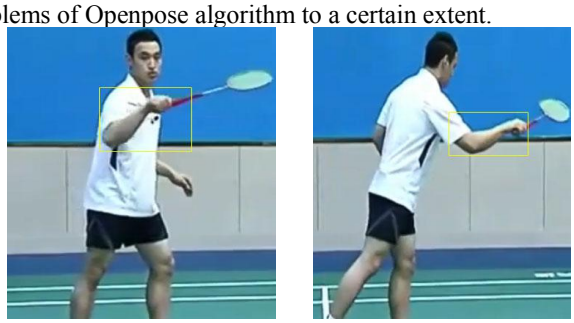


Figure 5. Tracking effect in badminton video

V. CONCLUSIONS

This paper chooses the video under the monitoring perspective of badminton match as the analysis object, and the real-time semantic analysis of badminton match includes two aspects: on the one hand, the semantic analysis of athletes' sports information, including sports data and action categories; on the other hand, the analysis of badminton movement track and location. Through the research and implementation of this semantic analysis technology, the movement in the badminton match field is tracked, recorded and analyzed, and data feedback and guidance are carried out in real time. This combination of online semantic analysis and offline competition provides the referee with teaching aids and the audience with fresh experience of watching the game, which is of great research significance.

REFERENCES

[1] Mezaris V, Kompatsiaris I , Strintzis M G . Video Object Segmentation Using Bayes-Based Temporal Tracking and Trajectory-Based Region Merging. IEEE Transactions on Circuits &

[2] Yang-Yang G , Dong-Jian H , Cong L . Target tracking and 3D trajectory acquisition of cabbage butterfly based on the KCF-BS algorithm. Scientific Reports, 2018, 8(1):9622-9634

[3] Kim W , Moon S W , Lee J , et al. Multiple player tracking in soccer videos: an adaptive multiscale sampling approach. Multimedia Systems, 2018, 24(6):611-623