WILEY | Hindawi

Research Article

Sports Video Motion Direction Detection and Target Tracking Algorithm Based on Convolutional Neural Network

Long Liu (ID)

Chongqing Preschool Education College, WanZhou 404100, Chongqing, China

Correspondence should be addressed to Long Liu; 20221005@hncj.edu.cn

In order to effectively detect and monitor athletes and record various motion data of targets, the study suggests a study of target tracking algorithms to detect the direction of motion video sports movement based on the neural network. A class of feedforward neural networks with convolutional computation and deep structure is one of the representative algorithms of deep learning. Firstly, the athlete image is obtained from the video frame; combined with the nonathlete image to construct the training set, use the bootstrapping algorithm to train the convolutional neural network classifier. In the case of input picture frames, pyramids of different scales are then constructed by subsampling, and the location of many candidate athletes is detected by a neural network of disruption. Finally, these centers calculate the center of gravity of the athletes, find the athlete to represent the candidate, and determine the location of the final athlete through a local search process. The results of the experiment show that the proposed scheme of 6000 frames in the two game videos is compared with the AdaBoost scheme, and the detection rate of the proposed scheme is 75.41% to calculate the average detection accuracy and false alarm speed of all players. The detection rate is higher than the AdaBoost scheme. Therefore, this scheme has a high detection rate and low false positives.

1. Introduction

Moving target tracking is a core subject in the field of computer vision. Its core idea is to capture moving targets quickly and accurately by comprehensively using technical means such as image processing and video analysis [1]. In recent years, with the continuous improvement of science and technology, the technology of moving target detection and tracking has also become mature. It has wide application prospects in medical research, traffic monitoring, passenger flow statistics, astronomical observation, visual monitoring, sports, and other fields. Monitoring dynamic scenes through cameras has long been widely used in all aspects of social life. Safety monitoring of public and vital facilities to control traffic on cities and highways, from the detection of military targets to intelligent weapons, cameras play a very important role as an extension of human vision [2, 3]. In sports videos, however, the athletes' colors and backgrounds are similar, and athletes can block each other out. The uniqueness of sports video poses significant challenges to object detection

and tracking technology. Mobile target tracking technology plays an important role in the field of sports video analysis. By tracking athletes in real time, we can analyze athletes' motion trajectory and judge the standardization of their actions. For example, in the diving competition, through the analysis of the athletes' diving action track, judge whether the athletes' take-off, somersault, entering the water, and other actions are correct and consistent. In weightlifting training, we can help athletes analyze the essentials of movement by tracking the movement track of barbell. Therefore, the purpose of this paper is to devise a plan to enable objective detection and tracking of heritage devices and to establish a simulation system to verify the accuracy of the algorithm. The goal of the motion detection and tracking system is to gain experience in digital imaging, modeling, computer vision, and other technologies (see Figure 1). This system can be widely used in related fields such as traffic control, astronomical observations, biomedical research, passenger traffic statistics, and sports [4–6]. In the analysis of sports video, moving target detection and
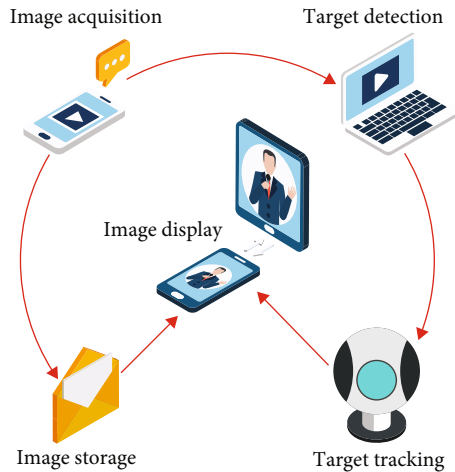
Figure 1: Sports video motion direction detection and tracking system based on convolutional neural network.

tracking technology has played an important role, which is convenient to correct the subtle movement differences that cannot be detected by human eyes in training or competition, thereby improving the training and competition results of the athletes. Therefore, together with the research topics of this paper, the research and application of technology in video game equipment were researched and discussed, combined with the research topic of this article, which will be of great theoretical importance and will have a positive impact on other areas based on mobile target tracking practical value [7, 8]. Detection and removal of moving objects in video games: detection and removal of moving objects are important for tracking of the target later, and the quality of the deleted image will be directly affected.

2. Literature Review

As a comprehensive application technology that plays an extremely important role in many fields, the research of video tracking theory started earlier. The research and application of multitarget detection, recognition, and tracking technology have been paid great attention [9]. Gardini and others proposed a color-based particle filter tracking algorithm, which uses the color histogram as the feature to track the moving target. Color histogram has the advantages of stable features, antipartial occlusion, simple calculation method, and small amount of calculation. The disadvantage is that when the distribution of background color is similar to that of target color, it is easy to mistake the background as a moving target. In particular, when the size of the tracked target is small, it is difficult to judge the exact position of the target according to the color histogram, and the convolution neural network depends on the number of sample points. When there are few samples, the accuracy of the algorithm will be reduced [10]. Wang and others believe that because the moving speed of the moving target is fast in sports video and the moving speed often changes greatly, it is difficult for the general motion model to accurately predict the approximate position of the moving target [11]. Kumar and others proposed a novel tracking algorithm based on target con-

tour, which uses the optical flow method to track the target contour. However, the optical flow method has complex calculation, low accuracy, and poor anti-interference and is vulnerable to noise [12]. Shang and others proposed a kernel-based mean shift algorithm, which has low computational complexity and high precision. By continuously calculating the mean shift vector, the search position can be updated iteratively until it converges to the optimal matching point. However, due to the limitation of convergence, the algorithm can achieve good results only if the difference between the predicted position and the real position of the target is small [13]. Véstias introduced a more general moving object detection and event recognition system. Objects are found by detecting interframe image changes, and a prediction and nearest neighbor matching technology is used in tracking [14]. Hidayat and others introduced a visual monitoring system. It uses multiple cooperative cameras to continuously track people and vehicles in complex environments and analyzes target categories and behaviors [15]. Gan and others proposed many algorithms for target tracking. According to the types of tracking targets, they can be divided into two categories: rigid object tracking and nonrigid object tracking. According to the number of targets, it can be divided into single target tracking and multitarget tracking [16]. Li and others estimated the motion of the object by calculating the motion of the brightness of the moving object table. In general, the motion of the object corresponds to the motion of the optical flow. Therefore, the relative motion of the object relative to the background can be obtained by calculating the optical flow field on the surface of the object. However, in practical application, due to the complexity of optical flow calculation and inaccurate estimation, it is less used [17].

Based on this research, this paper presents a sports video guidance research and goal tracking algorithm based on neural network connectivity. The bootstrapping algorithm is used to train the convolutional neural network classifier. For the input detection image frame, multiple candidate athlete positions are detected by convolutional neural network, and then, the candidate athlete positions are fused to determine the final athlete position. Experiments are carried out on some football game videos. Compared with the AdaBoost algorithm, the planning strategy achieves the best performance of the detection rate and alarm, and the search is faster.

3. Research Methods

3.1. Convolutional Neural Network Architecture. The convolution neural network is composed of six different types of convolution layers, as shown in Figure 2. The input layer receives the gray image of 21×43, and the C1 layer convolutes the input image using a 5×5 acceptance domain. This layer consists of 4 feature maps that share the receiving area and the deviation. Layer S1 performs subsampling and local averaging operations on the map, creating four feature maps [18]. Subsampling reduces both input dimensions and improves image translation, scale, and deformation stability. In addition, the map output of hybrid functions combines
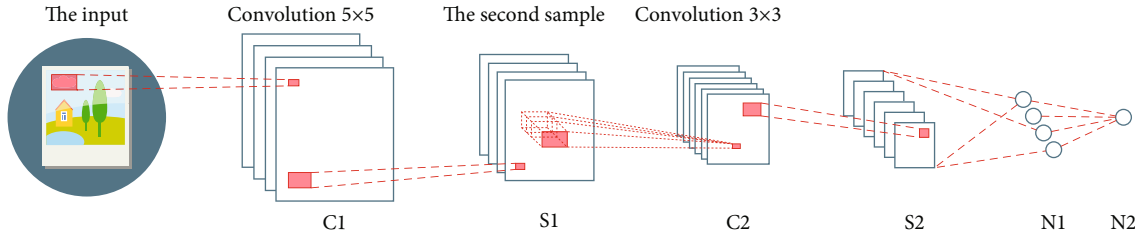
FIGURE 2: Basic structure of convolutional neural network.

different functions, which helps to extract more complex information. Layer C2 is not fully connected to layer S1, and the output map of layer S1 is converted to a 3×3 receiving area to create 14 feature maps. Layer S2 has the same function as layer S1 and consists of 14 feature maps. The role of layers N1 and N2 is to sort after the front section has been disassembled and the input size reduced. The output of a neuron in the N2 layer determines whether the input image is an athlete or not an athlete; -1 is not an athlete, and +1 is an athlete. In addition, for the training of network, this paper adopts the classical back-propagation algorithm with improved momentum method.

3.2. Proposed Athlete Detection Scheme

3.2.1. Training of Convolutional Neural Networks.
For the training of the network, the bootstrapping strategy is adopted. It is a self-expanding method that initializes a learner with seed information and seed templates and expands new knowledge and improves learning performance by automatically learning the training set. Apply a neural network on a sample set containing nonathlete images and iteratively enhance the negative training sample set based on the resulting false positives. The algorithm steps are shown in Table 1.

Bootstrapping algorithm mainly has the following steps. (1) Establish a test data set composed of athlete images (positive samples) and nonathlete images (negative samples). The test set remains unchanged in the bootstrapping iteration. On the contrary, the training set needs to be updated continuously. (2) For the neurons in N1 and N2 layers, a back-propagation algorithm with increasing momentum term is used to train the network. In the iteration, ThrFa gradually decreases to 0, which can avoid the redundancy of some training sets. (3) Select the samples whose false alarm result of network classification is greater than ThrFa, generate a new model, and add it to the negative sample training set so that the network will focus on the decision boundary of current athlete classification in the next iteration. After 6 iterations, the learning process stops when the number of false positives remains approximately constant.

3.2.2. Detection of Athletes.
This paper is based on the trained convolution neural network to detect athletes. The specific process is mainly divided into the following five steps.

Step 1. In order to detect athletes with multiscale size, repeat the secondary sampling operation with a factor of 1.2 on the input image to generate a pyramid composed of images with different sizes and scales.

Step 2. For each image of the pyramid, complete convolution is carried out through the convolution neural network to obtain an image containing the output results of the network. The positive pixels in the output image are the detected candidate athlete positions.

Step 3. This paper observes that real athlete images usually give a continuous scale value of positive response, while nonathlete images will not occur. In order to eliminate false positives to real athletes, this paper determines the distribution based on the volume of the positive solution (the positive value of the positive solution) in the local pyramid. If its volume is more than that of the original ThrVol, the athlete is classified as an athlete; otherwise, he is a nonathlete [19, 20].

3.3. Overview of Object Detection Methods

3.3.1. Image Preprocessing Method.
For the collected original pictures, due to noise, light, and other reasons, which often cannot be directly used for tracking and detection, so first we need to carry out the relevant preprocessing work of the original image. Preprocessing includes image processing, file encoding and transmission, edge sharpening, and more. Preprocessing not only effectively removes images and improves image quality and sharpness but also ensures good processing for target processing, such as targeted detection, extraction, and timely monitoring of target time. It is more suitable for computer analysis, image comprehension, and recognition. In sports video target detection and monitoring system, many conventional algorithms related to image processing are usually used. Generally speaking, these auxiliary technologies are often used before the core processing technology, and their purpose is to improve the performance of the system. The image preprocessing technology used in this paper mainly includes image enhancement, ordinary filtering, and morphological filtering.

(1) Gray Processing. Color images are generally divided into three types: black and white, grayscale, and color. In general engineering applications, it is often necessary to convert color images into grayscale shapes to solve problems. Digital video recordings captured by digital cameras are all color

1. Randomly selected 50 positive and 50 negative samples from the initial training to develop the test procedure. It will be used to select the recommended weight in grades 3 and 8

2. Set momentum parameter BIter = 0 and false alarm threshold ThrFa = 0.7

3. Iterate the training network 50 times, and use the same number of positive and negative samples in each iteration. Set BIter = BIter + 1

4. From 100 video frames, collect samples whose detection false-positive results exceed ThrFa, and collect up to 3000 new samples

5. Add the newly collected samples to the negative sample training set

6. If ThrFa ≥ 0.2, set ThrFa = ThrFa − 0.2

7. If BIter < 6, go to step 3

8. Iterate 50 times and exit

images; in order to facilitate fast processing, it is necessary to replace the printed images' color in gray. The process of converting color images to a gray image is called grayscale processing. The description of the grayscale image, like the color image, also shows the total and local distribution and the characteristics of chromaticity and brightness of the whole image. Typically, each pixel of the color image is represented by 3 bytes, each byte corresponds to the brightness of the RGB component, and each pixel of the converted image is represented by one byte. The higher the brightness, the brighter; the lower the value, the darker the light. The relationship between conversions usually uses the following model:

$$\text{gray}(i, j) = 0.11r(i, j) + 0.48g(i, y) + 0.2b(i, j). \quad (1)$$

Gray conversion can also take the maximum, minimum, or arithmetic average of the three components, respectively. Gray processing first reads the image and copies it to the memory and then makes each color component equal and equal to $\text{gray}(i, j)$, which completes the process of converting the original color image into gray image.

(2) Image Enhancement. The purpose of image enhancement is to enhance the information that users are interested in the image, such as edge and contour, expand the difference between different object features in the image, and provide a good foundation for the extraction of image information and the application of other analysis technologies. The general formula for the conversion of grays is shown in

$$s = T(r). \quad (2)$$

In the formula, r and s represent the pixel values before and after processing, respectively, and T is a transformation that maps from the original definition domain $[r_0, r_k]$ to the new definition domain $[s_0, s_k]$.

Different definitions of T can get different transformation results. The commonly used grayscale transformations include linear inversion transformation, logarithmic transformation, contrast stretching, and histogram equalization. The function formula of contrast stretching is (3), where e is the parameter given by the control slope and M is the

mean value of pixel gray. The output of the narrow frame is a high-contrast image.

$$s = T(r) = \frac{1}{1 + (m/r)^E}. \quad (3)$$

Histogram averaging changes the input gray level according to formula $(x \cdot x)$ to obtain the output gray level s, as shown in

$$s = T(r) = \int_0 p_r(w)dw. \quad (4)$$

In the formula, $p_r(w)$ is a function of the probability density of the gray level in a given figure and w is the dummy variable of the integral [21]. Then, as shown in Equation (5), the probability density function of the output gray level is uniform:

$$p_s(s) = \begin{cases} 1, 0 \leq s \leq 1, \\ 0. \end{cases} \quad (5)$$

After histogram equalization, the gray level of the image is more balanced, and the final result is an image with extended dynamic fan Tian, which has high contrast.

3.3.2. Commonly Used Moving Target Detection Methods. Detecting moving objects in the sequence of images is a difficult and very important field of study. In general, object detection in sports video mainly identifies and analyzes moving objects in the video stream and filters out moving objects in the image from the scene. Commonly used detection methods are the range difference method, the background removal method, the statistical method, and the optical flow method.

(1) Frame Difference Method. The range difference method is an algorithm that uses sequential frame image differences in a video sequence for target detection and resolution. This is a very common method. Threshold processing plays a key role in the application of frame difference method, because if the threshold is too low, it will suppress the effective changes in the image. The selection of threshold usually depends on the specific external environmental conditions
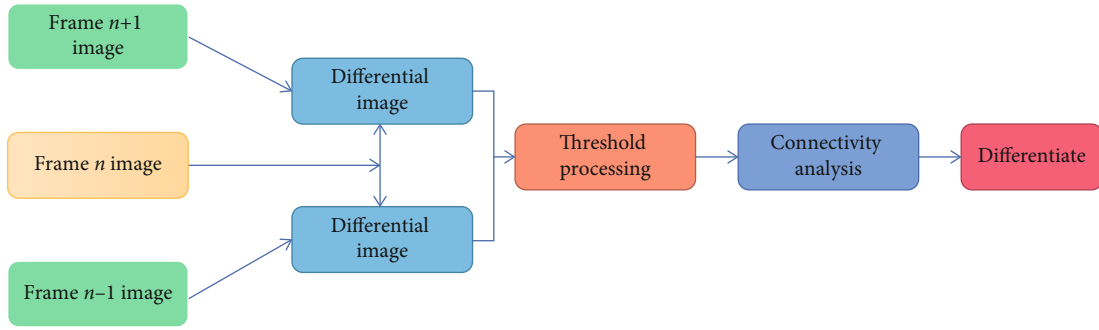
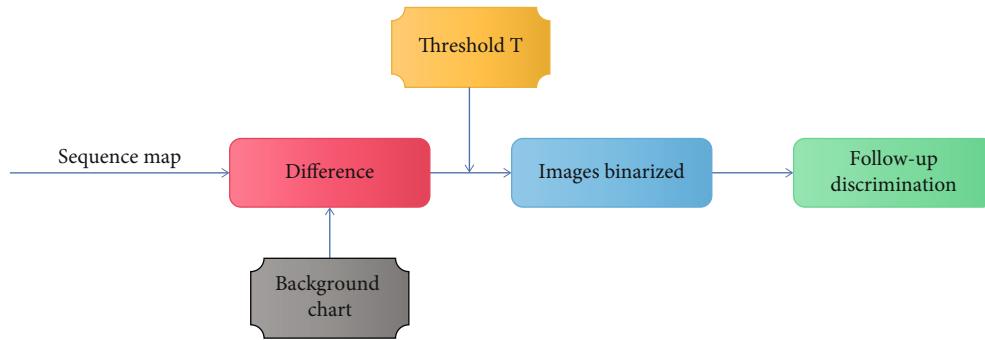FIGURE 3: Schematic diagram of the frame difference method.

FIGURE 4: Background subtraction method.

such as scene and camera. The selection of threshold can choose either global threshold or local threshold, because the noise caused by the image under different illumination is not necessarily the same, so the use of local threshold can better suppress the noise.

The algorithm of frame difference method is simple and does not consider the update of background, and its shortcomings are also very obvious. The number of frames taken by the algorithm is high, and the moving speed of the target is also required. If the target moves too fast and the selection time interval is long, there will be no coverage area between frames, resulting in the inability to segment the moving target. On the contrary, Figure 3 is a schematic diagram of the frame difference method.

(2) Cut the Background. The background removal method is also a common algorithm for detecting moving targets. Its principle is to exclude algorithms that use current images and background images to achieve moving targets (see Figure 4), used to remove the current frame and background pattern. If the background pattern is chosen correctly, moving objects can be segmented more accurately. The background subtraction method is generally based on a fixed camera. In principle, if the background is still, the pixels of the video image with moving targets other than the moving targets should be unchanged, and only the moving target area changes. How to get this invariant region to meet the dynamic changes of the scene is a difficulty in the background subtraction method. Considering that background subtraction is a changing process, it is necessary to update

the background model at any time according to different situations, that is, to increase the adaptability of the algorithm itself. Background subtraction is the process of subtracting each frame in the image sequence with a fixed background model. Its mathematical expression is shown in

$$R(i, j) = F(i, j) - G(i, j), \qquad (6)$$

where $R(i, j)$ is the moving target to be detected; $F(i, j)$ is a video sequence image; $G(i, j)$ is the background model image.

(3) Optical Flow Method. The optical flow method analyzes the motion field of each point in the sequence image to find out the motion of the corresponding point on the image plane caused by spatial motion. Optical flow method usually assumes that the interval between adjacent times is very narrow, which is generally considered to be within tens of milliseconds, so the difference between images at adjacent times is also very small. The optical flow method does not need to process the image and extract its eigenvalues first, but directly process the image itself.

3.4. Multitarget Tracking Algorithm Based on Camera Motion Estimation

3.4.1. Global Motion Estimation. Global motion is usually caused by the movement of the camera. If the camera moves during shooting and the objects in the frame have their own motion, then the background and foreground have their

own motion. In the video sequence, the sound of the background is caused by the sound of the camera that is called global motion. The goal of global forecasting is to find the right camera sound system that allows the world to move through the video on a regular basis. In video segmentation of moving objects, you can first calculate global motion, then compensate for camera movement between calculated frames to align the background between frames, and then separate the front object and background according to the motion zone information. When generating panorama, the correlation of corresponding pixels between frames is obtained by global motion estimation. Then, the panorama can be obtained by stitching the adjacent frames according to the motion parameters. Encoding is the use of panoramas to predict and compensate, which greatly improves the compression ratio. Therefore, the analysis of the law of motion of the camera, which leads to a change in the image, or the analysis of the motion of the front object, is the basis for the analysis of the motion of the world. Methods for estimating global motion parameters are generally divided into differential methods and point-to-point methods. In this paper, the six-parameter affine model is used to model the camera motion causing the scene change between frames, and the differential method is used to solve the global motion parameters. Since the above conditions can be met between adjacent frames during video capture, such a model can reasonably depict the movement of the camera between adjacent frames.

The motion of the earth on the background due to the motion of the chamber can be expressed by a model of affine motion with 6 parameters, as shown in

$$\begin{cases} x_i = ax_i' + by_i' + e, \\ y_i = cx_i' + dy_i' + f. \end{cases} \tag{7}$$

Among them, $p = (x_i, y_i)$ is the coordinate of the current frame I_k, $p'' = (x_i', y_i')$ is the adjacent frame, the coordinates of the point corresponding to P in I_k.

3.4.2. Camera Model and Camera Calibration.
The camera design simplifies and approximates the geometry of the optical image. Camera design is usually defined by a number of parameters called camera parameters, and the process of resolving camera parameters is called camera adjustment. The perforated model is the most suitable model for the camera. It defines the descriptive process as the central process of perspective planning. The intersection of the line connecting the point on the scene with the optical center and the plane of the image is the point of projection of the point on the image. Perspective is characterized by "near is big, far is small." Also, the points on the line do not change the ratio of the intersections during the projection. Figure 5 shows the projection process of a camera simulated by computer graphics. We call the joint location as the camera joint, and the joint design with the camera as the center location and as the direction of the camera is called the control camera. The image control system is an integrated system created by two-dimensional images, which generally acts as
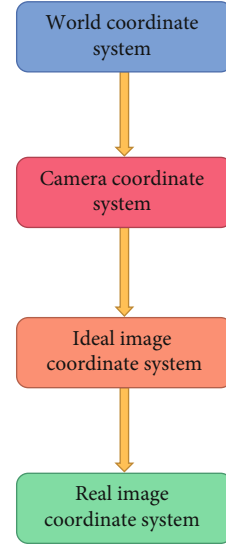

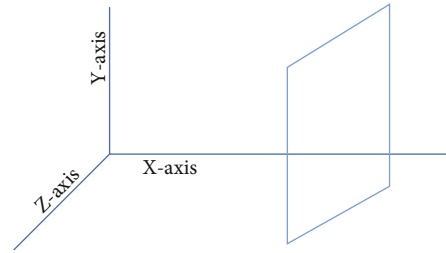FIGURE 5: The process of projecting a point in the world coordinate system to an image.

FIGURE 6: Pinhole camera model.

a camera control system. Figure 6 is a schematic diagram of the model of the pinhole camera.

3.4.3. Sports Vision's Many Target Trajectory Tracking Algorithm.
Multitarget monitoring is the focus of current computer visual research, especially human tracking, which is the current research hotspot. Current multitarget control algorithms are roughly divided into two types: model-based multitarget control systems, a multitarget monitoring system based on the integration of information from multiple sources. The model-based multitarget control algorithm mainly uses multitarget motion models to create a multitarget motion model using the relationships between tracks and then uses the corresponding tracking algorithm to search for the state position space to obtain the target position. It is mainly used to track people. A multisource target tracking algorithm is usually used to melt information obtained from multiple sensors and then uses a neural network or latent Markov model to integrate the information. This type of algorithm is mainly used in radar signal processing and other fields.

This article uses rotating neural networks and camera motion algorithms to analyze football and hockey videos, track athletes' treadmills, obtain athletes' movement calculations and movement speeds, and assist coaches in tactical
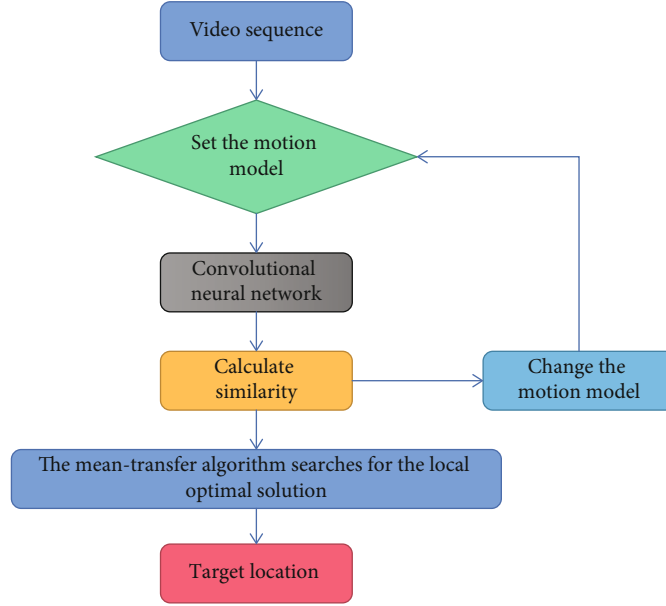
FIGURE 7: Algorithm flowchart.

analysis. Traditional multitarget tracking algorithms are usually based on a static background, and this type of algorithm cannot be useful because it is not possible to obtain real-time target movement information because it is only possible to obtain the target speed and trajectory compared to the camera information for coaches. The flow of algorithms in this article is shown below.

(1) Use the camera calibration algorithm to obtain the mapping relationship between the site model and the first video image, as shown in

$$\begin{cases} x' = \dfrac{a_{11}x + a_{12}y + a_{13}}{a_{31}x + a_{32}y + a_{33}}, \\ y' = \dfrac{a_{21}x + a_{22}y + a_{23}}{a_{31}x + a_{32}y + a_{33}}, \end{cases} \quad (8)$$

where point (x', y') is the coordinate on the site model and point (x, y) is the coordinate on the first video image

(2) Using convolution neural network and mean shift hybrid tracking algorithm, the coordinate point (x_k'', y_k'') of the player on the current video frame is obtained

(3) $(x_i, y_i)^T$ is the position of a pixel point in the current frame image and $(x_i', y_i')^T$ is the position of the point in the previous frame image. The relationship between the two is shown in

$$\begin{pmatrix} x_i' \\ y_i' \end{pmatrix} = A \begin{pmatrix} x_i \\ y_i \end{pmatrix} + T, \quad (9)$$

where

$$A = \begin{pmatrix} a_2 & a_3 \\ a_4 & a_5 \end{pmatrix} \quad (10)$$

represents scaling, rotation, and stretching motion; $T = (a_0, a_1)^T$ represents translational motion. The camera motion parameter

$$\begin{pmatrix} a_{k0} & a_{k2} & a_{k3} \\ a_{k1} & a_{k4} & a_{k5} \end{pmatrix} \quad (11)$$

is obtained by using the global motion estimation algorithm

(4) Solve that the coordinate point (x_k'', y_k'') of the tracked target on the current frame corresponds to the coordinate point $(\tilde{x}_k, \tilde{y}_k)$ under the image coordinate system of the first frame, as shown in

$$(x_k^n, y_k^n)^T = \prod_{j-1}^{k} \begin{pmatrix} a_{j0} & a_{j2} & a_{j3} \\ a_{j1} & a_{j4} & a_{j5} \end{pmatrix} \begin{bmatrix} 1 \\ \tilde{x}_k \\ \tilde{y}_k \end{bmatrix}, \quad (12)$$

where

$$A = \prod_{j-1}^{k} \begin{pmatrix} a_{j0} & a_{j2} & a_{j3} \\ a_{j1} & a_{j4} & a_{j5} \end{pmatrix} = \begin{pmatrix} a_0 & a_2 & a_3 \\ a_1 & a_4 & a_5 \end{pmatrix},$$

$$\begin{pmatrix} a_{j0} & a_{j2} & a_{j3} \\ a_{j1} & a_{j4} & a_{j5} \end{pmatrix} \quad (13)$$

TABLE 2: Technical parameters.

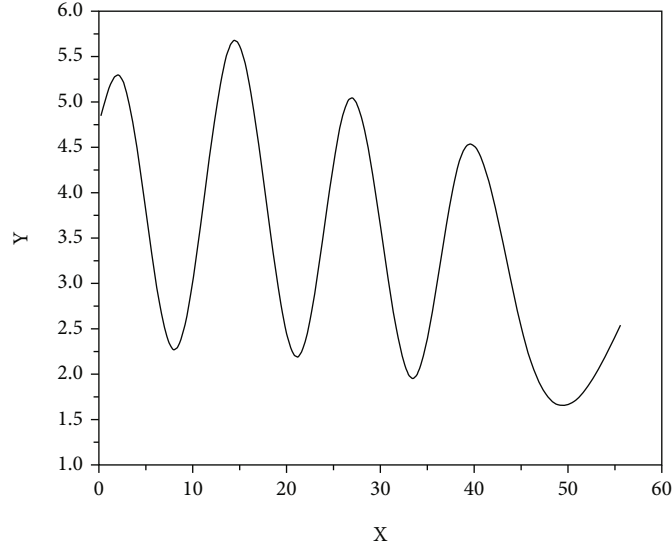| Test video	Resolving power	Number of frames	Number of color blocks of color histogram	Number of samples of convolutional neural network	Mean shift average iteration	Calculation time (MS)
Table tennis	251×177	74	$3 * 3 * 3$	100	5	2431
Horizontal bar	450×270	76	$7 * 7 * 3$	100	3	4063
6142 football	241×177	70	$7 * 7 * 7$	300	4	6142
Weightlifting	610×465	67	$7 * 7 * 7$	300	3	11820

FIGURE 8: Schematic diagram of table tennis motion tracking trajectory.

TABLE 3: Comparison of the average detection rate and false alarm rate of the two schemes for each team's player detection.

Team	Paper scheme		AdaBoost scheme	
	Detection rate (%)	False-positive rate (%)	Detection rate (%)	False-positive rate (%)
England	76.81	1.10	63.61	7.68
Portugal	83.14	1.04	88.55	1.15
France	87.77	0.41	83.27	1.14
Italy	73.36	1.83	72.40	1.47
Average value	80.07	1.08	74.30	1.60

is the global motion estimation parameter from frame $J - 1$ to frame J of the video. The flowchart of this algorithm is shown in Figure 7. Table 2 is the technical data obtained during the testing of this document, and Figure 8 is a schematic diagram of the table tennis track control

4. Discussion of Results

The proposed method is compared with the detection method based on AdaBoost algorithm. The video set used for training and testing is recorded by a fixed position mobile camera. In the video frame, the player position of a specific party is manually marked and represented by a rect-

TABLE 4: Comparison of the average detection rate and false alarm rate of the two schemes for all player detection.

	Detection rate (%)	False-positive rate (%)
Paper scheme	75.41%	1.62%
AdaBoost scheme	70.02%	1.34%

angular box. Then, these athlete images are extracted with a size of 12×24 pixels [22].

4.1. Detection of Specific Athletes. In the first experiment, the specific players in the game video were detected. Extract video samples from two matches of FIFA World Cup. Among them, each team wears different colors of team
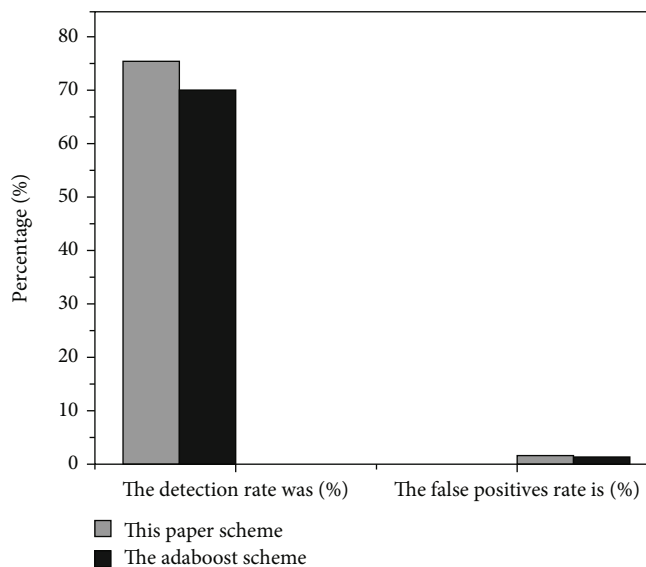
FIGURE 9: Comparison of the average detection rate and false alarm rate for all players detected by the two schemes.

uniforms, so as to form a different contrast compared with the background. For each team, 250 samples containing negative samples were extracted to form a training set, and four independent detectors corresponding to four teams were trained and tested on each team. Compare this scheme with AdaBoost scheme on 6000 frames of the above two game videos, and calculate the average accuracy and false-positive rate of player detection of each team. The results are shown in Table 3. Among them, the accuracy rate is the proportion of the number of athletes detected as the team in one frame to the number of all athletes of the team. The false-positive rate is the proportion of the number of athletes in the team to the number of nonathletes in the team.

4.2. Testing of All Athletes. In the second experiment, all players on the game video were detected. In this paper, 450 samples are extracted from the above two games as training sets to train the detectors corresponding to the two games. Compare this scheme with AdaBoost scheme on 6000 frames of the above two game videos (see Table 4), and calculate the average accuracy and false-positive rate of all players. The detection rate of this scheme is 75.41%, which is much higher than that of AdaBoost scheme. The results are shown in Figure 9.

5. Conclusion

The complexity of sports itself brings many difficulties to the actual detection and tracking of moving objects. In order to detect and track athletes effectively, this paper improves the commonly used single tracking algorithm and further improves the effect of sports target detection and tracking. To identify the benefits of an algorithm, this paper uses Matlab to simulate and provide examples of discovery and tracking effects. The main research results are as follows. Detection and removal of moving objects in video games:

detection and removal of moving objects are important for tracking of the target later, and the quality of the deleted image will be directly affected, effect after tracking. In view of the difficult problem in video game, this paper compares the advantages and disadvantages of different processes, background deletion process and streamer process, and also compares medium, medium, morphological filtering, and other filtration algorithms. Video games move the extraction target based on convolutional neural network with experimental quality.

In this paper, some main problems related to sports video moving target tracking technology are studied, and the corresponding research results are obtained. However, the relevant video processing technology still needs further research in the following aspects. At present, video-based target tracking algorithms can only deal with the visual information from a single angle and cannot obtain the omnidirectional information of the whole tracked target as a whole. Therefore, this kind of algorithms are often difficult to achieve good results when dealing with blocking, heavy prosperity, and other phenomena. If the visual information from multiple angles can be fused, the stereo feature model of the tracked target can be established, and the correlation between cameras can be used to track the motion of the target; the robustness and stability of the algorithm will be greatly improved. Therefore, the future multitarget tracking algorithm will develop in the direction of multisource information fusion.

Data Availability

No data were used to support this study.

Conflicts of Interest

The author declares that there are no conflicts of interest regarding the publication of this article.

Acknowledgments

References

[1] A. N. Gorban, E. M. Mirkes, and I. Y. Tukin, "How deep should be the depth of convolutional neural networks: a backyard dog case study," *Cognitive Computation*, vol. 12, no. 2, pp. 388–397, 2020.

[2] Z. Huang, J. Zhu, J. Lei, X. Li, and F. Tian, "Tool wear monitoring with vibration signals based on short-time Fourier transform and deep convolutional neural network in milling," *Mathematical Problems in Engineering*, vol. 2021, Article ID 9976939, 14 pages, 2021.

[3] H. Ali, G. Mubeen, T. S. Ali, Z. Tehseen, and A. Waqas, "High efficiency video coding (hevc)-based surgical telementoring system using shallow convolutional neural network," *Journal of Digital Imaging*, vol. 32, no. 6, pp. 1027–1043, 2021.

[4] S. Qin, R. V. Lehn, V. Zavala, and T. Jin, "Predicting critical micelle concentrations for surfactants using graph convolutional neural networks," *The Journal of Physical Chemistry B*, vol. 125, no. 37, pp. 10610–10620, 2021.

[5] K. C. Burak, M. K. Baykan, and H. Uuz, "A new deep convolutional neural network model for classifying breast cancer histopathological images and the hyperparameter optimisation of the proposed model," *The Journal of Supercomputing*, vol. 77, no. 3, pp. 1–17, 2021.

[6] Y. Cheng, K. Hu, J. Wu, H. Zhu, and X. Shao, "A convolutional neural network based degradation indicator construction and health prognosis using bidirectional long short-term memory network for rolling bearings," *Advanced Engineering Informatics*, vol. 48, no. 1, pp. 101247–101251, 2021.

[7] K. Marek, E. Micha, S. Marcin, K. Józef, and M. Roman, "Cell nuclei segmentation in cytological images using convolutional neural network and seeded watershed algorithm," *Journal of Digital Imaging*, vol. 33, no. 1, pp. 231–242, 2021.

[8] W. Cotrim, L. B. Felix, V. Minim, R. C. Campos, and L. A. Minim, "Development of a hybrid system based on convolutional neural networks and support vector machines for recognition and tracking color changes in food during thermal processing," *Chemical Engineering Science*, vol. 240, no. 4, pp. 116679–116683, 2021.

[9] S. Baroud, S. Chokri, S. Belhaous, and M. Mestari, "A brief review of graph convolutional neural network based learning for classifying remote sensing images," *Procedia Computer Science*, vol. 191, no. 1, pp. 349–354, 2021.

[10] E. Gardini, M. J. Ferrarotti, A. Cavalli, and S. Decherchi, "Using principal paths to walk through music and visual art style spaces induced by convolutional neural networks," *Cognitive Computation*, vol. 13, no. 2, pp. 570–582, 2021.

[11] Y. Wang, J. Peng, and Z. Jia, "Brain tumor segmentation via c-dense convolutional neural network," *Progress in Artificial Intelligence*, vol. 10, no. 2, pp. 147–156, 2021.

[12] A. Kumar, A. R. Tripathi, S. C. Satapathy, and Y. D. Zhang, "Sars-net: covid-19 detection from chest X-rays by combining graph convolutional network and convolutional neural network," *Pattern Recognition*, vol. 122, no. 11, article 108255, 2021.

[13] R. Shang, Y. Meng, W. Zhang, F. Shang, and S. Yang, "Graph convolutional neural networks with geometric and discrimination information," *Engineering Applications of Artificial Intelligence*, vol. 104, no. 1, article 104364, 2021.

[14] M. Véstias, "Efficient design of pruned convolutional neural networks on fpga," *Journal of Signal Processing Systems*, vol. 93, no. 5, pp. 531–544, 2021.

[15] A. A. Hidayat, T. W. Cenggoro, and B. Pardamean, "Convolutional neural networks for scops owl sound classification," *Procedia Computer Science*, vol. 179, no. 4, pp. 81–87, 2021.

[16] H. Gan, M. Ou, F. Zhao, C. Xu, and Y. Xue, "Automated piglet tracking using a single convolutional neural network," *Biosystems Engineering*, vol. 205, no. 1, pp. 48–63, 2021.

[17] J. Li and D. Liu, "Information bottleneck theory on convolutional neural networks," *Neural Processing Letters*, vol. 53, no. 2, pp. 1385–1400, 2021.

[18] C. P. George, "Convolutional neural networks: alternate drivers' visual perception," *IEEE Potentials*, vol. 39, no. 1, pp. 19–24, 2020.

[19] J. Wang, T. T. Hormel, Q. You, Y. Guo, and Y. Jia, "Robust non-perfusion area detection in three retinal plexuses using convolutional neural network in oct angiography," *Biomedical Optics Express*, vol. 11, no. 1, pp. 330–345, 2020.

[20] Y. Xu, H. Cui, B. Fan, B. Zou, and L. Wang, "Integrative model of CT imaging and clinical features using attentional multiview convolutional neural network (AM-CNN) for prediction of esophageal fistula in esophageal cancer," *International Journal of Radiation Oncology • Biology • Physics*, vol. 108, no. 3, pp. e637–e638, 2020.

[21] W. Walid, M. Awais, A. Ahmed, G. Masera, and M. Martina, "Real-time implementation of fast discriminative scale space tracking algorithm," *Journal of Real-Time Image Processing*, vol. 18, no. 6, pp. 2347–2360, 2021.

[22] Z. Cheng, Z. Yang, and Y. Gangui, "A novel frequency regulation strategy for a pv system based on the curtailment power-current curve tracking algorithm," *IEEE Access*, vol. 8, pp. 77701–77715, 2020.