Hindawi
Mobile Information Systems
Volume 2022, Article ID 4805993, 10 pages
https://doi.org/10.1155/2022/4805993

Research Article

A Sports Video Behavior Recognition Using Local Spatiotemporal Patterns

Yuanqiang Liu[1] **and Huaiguo Jing**[2]

[1]*Guangdong Engineering Polytechnic, Guang Dong 510520, China*
[2]*Guangdong University of Technology, Guang Dong 510006, China*

Correspondence should be addressed to Huaiguo Jing; jinghuaiguo1978@163.com

Received 9 March 2022; Revised 8 April 2022; Accepted 26 April 2022; Published 6 June 2022

Academic Editor: Mian Ahmad Jan

These days, many researchers are interested in sports video behavior recognition (SVBR), which has emerged as a core research domain for visual understanding and analysis of athlete performance. Despite significant results in simple scenarios, SVBR remains a difficult task due to numerous challenges such as independence, obstruction, and interclass visual appearance in real situations. Recognizing the sports video behavior is one of the crucial challenges in computer vision for sports video behavior data, with implementation in monitoring, mental illness diagnostics, video information extraction, and multitarget tracking. This paper provides a local pattern activity discrimination model for the detection and localization of active individuals in a video to address issues of multitarget tracking in video behavior recognition caused by mutual occlusion of targets and the complexity of ambient background. This work solves the problem of low recognition accuracy due to incomplete extraction of trajectories or overly complex backgrounds using a trajectory-based approach. Our model is applied to the video segments of 38 matches associated with goal events in Euro 2012, and an average accuracy of 91.3% is obtained. The experimental results verify the high accuracy and applicability of our method for the recognition of target object behavior in the videos.

1. Introduction

Recognition and analysis of human action [1] are two of the most active topics in computer vision, which are gaining attraction among researchers. This is because it has applications in video surveillance, video interpretation, retrieval, and human-computer interaction. But human action recognition is difficult due to factors such as high intraclass variance, scaling, occlusion, and disorder. On the other hand, behavior recognition encompasses a wide range of disciplines, including pattern recognition and computer vision, which are widely used in the fields of human movement analysis and security monitoring [2, 3]. The standard behavior recognition methods are classified into two categories: one is based on the target's trajectory, and the other is based on the target's features [4], and there are numerous research findings. In terms of trajectory-based behavior recognition, the early work of [5] used the trajectory information of five normalized nodes of the human head and torso to construct human behavior and then recognize it. Similarly, the authors of [6] employed support vector machines (SVM) to recognize three group rows based on motion trajectory and appearance information, while the early work of [7] used the analysis method of time/space interaction relationship and used player trajectory and ball trajectory information to construct interaction trajectories. They successfully identified tactical behaviors based on interaction trajectories. Besides, the work of [8] identified group behaviors by analyzing the interaction between individual trajectories.

Current sports video behavior recognition technologies have accomplished great results in the general visual recognition problem based on images [9], but there are very limited methods and data for video behavior recognition in scenes of the sports video. Precise identification of athlete behavior is a critical link in sports video analysis for a fresh information field. The existing approaches have accomplished great results in the common human body behavior assignment and in the physical training video, which will identify spectators and sportspeople at the same moment, making it impossible to differentiate athletes' objectives, which will interact with the successive video analysis. Simultaneously, sports video information with human body explanation is uncommon [10], and the cost of developing an effective model for the area of sports video is great. Sportsmen are a subset of the common human behavior detection process, so if the common human video behavior recognition model can be utilized to identify and detect sportspeople in sports videos, it will surely save a significant amount of money. As a result, dealing with these issues necessitates the use of an efficient technique.

The task of categorizing which action is being conducted in a series of frames, as well as localizing identification in both time and space, is known as spatiotemporal activity recognition [11]. In this technique, bounding boxes or masks can be used to illustrate the clustering. Because of the wide availability of computational services as well as recent technologies in convolutional techniques, there has been an increase in interest in this assignment in recent years. In terms of target-based behavior recognition, the work of [12] proposed a human behavior recognition method based on spatiotemporal posture features that can effectively recognize complex human action behaviors in the video. The work of [13] created visual information using spatiotemporal patterns, detected targets using background subtraction, represented the spatial distribution of targets in different regions using symbolic sequences, and then realized video behavior recognition. In this regard, the author of [14] proposed a method for recognizing human behavior based on low-level features and high-level semantics and obtained the final verdict human behavior category by fusing the two features' preclassification results. Earlier work by [15] combined the spatiotemporal characteristics of human behavior and proposed a spatiotemporal feature fusion deep learning network for human skeleton behavior recognition method with strong robustness to multiview skeletons.

Considering that the behavior recognition method based on trajectory information is susceptible to occlusion, background changes, and other factors [16, 17], this paper proposes a behavior recognition method based on local spatiotemporal patterns utilizing the 2D local regression kernel. The rest of the contributions to this paper are listed as follows:

(1) First, the improved local spatiotemporal regression kernel is used as a feature detector by detecting the active regions of motion. After that, they are used as feature words, and the target objects are labeled using a color information-based method. Consequently, the feature bag-of-words model is constructed.

(2) Secondly, the behavior is classified and recognized according to the location and behavior of the target in the field using the LibSVM pattern classifier.

(3) Finally, the method described in this paper is applied to a soccer match video to analyze all of the goals scored by Barcelona teams during the Euro 2012 [18] match. Furthermore, to identify the primary tactical behaviors used by the entire match and a single team based on the results of manual classification and labeling, we calculated the accuracy rate and tested the usefulness of this paper method.

The rest of this paper is designed as follows: Real-time athlete behavior recognition models in sports video are clarified in Section 2, experimental work and simulations are presented in Section 3, and finally, the conclusions are offered in Section 4.

2. Real-Time Athlete Behavior Recognition Models in Sports Video

There are strict laws regarding sports videography and formatting methods. The characteristics of the various shots, as well as the differences between them, are noticeable. The area of the tournament has the highest proportion in the round shot, and athletes are also focused on the game. As a result, the competition game area is identified first because once athletes have been identified, more precise sportspeople can be gained by eliminating the color of the field by the sports field's reliable color features. Because the field color still accounts for a portion of the image in the center lens, if the scene of far lens has been produced, the main area color is saved to eliminate the nonathletic area of the image that belongs to the competition ground.

Athlete detector is a classification that uses intermediate characteristic blocks to train and differentiate athletes by combining the feature and backgrounds of athletes in a variety of shots in sports videos. The intermediate neural network features frequently contain a wealth of information that plays a cohesive role throughout the neural network. We present in this paper a real-time athlete behavior detection model in sports footage based on a deep spatiotemporal residual convolutional neural network. To extract spatial and temporal features of athlete behaviors in sports videos, the model employs a deep spatiotemporal residual convolutional neural network. Furthermore, it builds a deep spatiotemporal residual convolutional neural network that can automatically learn athletes' spatiotemporal behavioral features from video data. The deep features obtained in this manner contain high-level behavioral information and are better suited for understanding athlete behavior [19, 20]. On this basis, the athlete's behavior is classified based on classification learning to achieve real-time recognition of athlete behavior. The structure of the constructed real-time athlete behavior recognition model based on a deep spatiotemporal residual convolutional neural network is shown in Figure 1.

2.1. Athlete Target Detection. Detecting and locating athletes' targets in sports videos is the key to realizing athletes' behavior recognition. This paper uses the video target detection algorithm based on deep learning to realize the real-time detection of athlete targets in sports videos [21]. The deep learning network structure of target feature

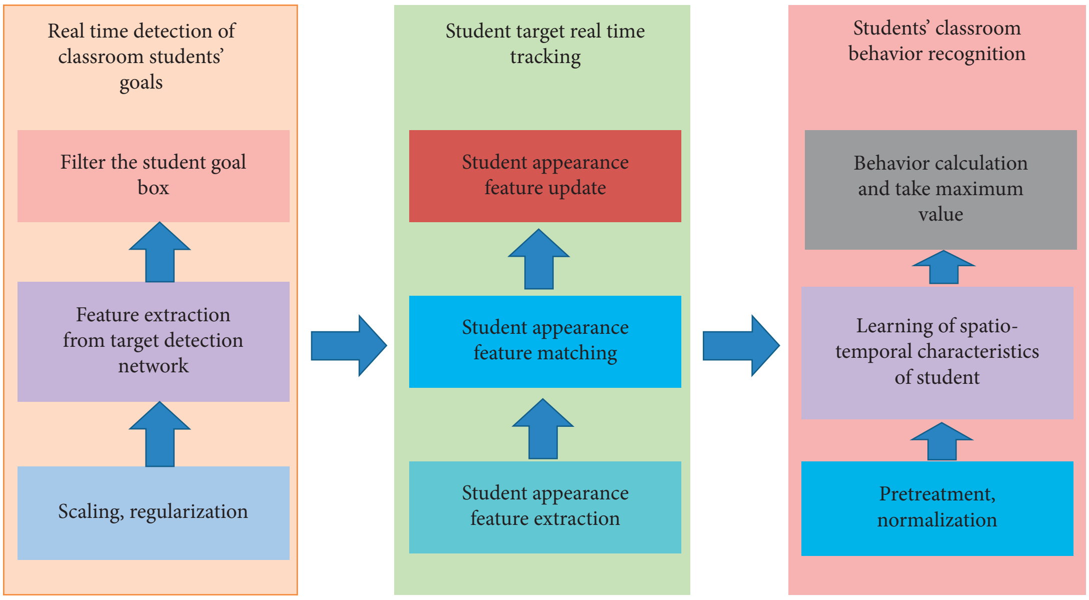
Real time detection of classroom students' goals
Filter the student goal box
Feature extraction from target detection network
Scaling, regularization
Student target real time tracking
Student appearance feature update
Student appearance feature matching
Student appearance feature extraction
Students' classroom behavior recognition
Behavior calculation and take maximum value
Learning of spatio-temporal characteristics of student
Pretreatment, normalization

Figure 1: Structure of athlete behavior recognition model based on spatiotemporal depth residual convolution neural network.

learning and detection includes a deep convolution neural network for extracting target features in sports video frames, a feature fusion network integrating multilayer features, and a convolution network for target detection. The structure diagram of classroom athlete target real-time detection network based on deep learning is shown in Figure 2.

The convolution process of each layer of the deep convolutional neural network for target feature learning in sports video frames is shown in the following:

$$h_l(X) = (X * W + b) \otimes \sigma(X^* V + c), \qquad (1)$$

where $x \in R^{N\times m}$ is the input of each layer, $w \in R^{k\times m\times n\times m}$ represents the weight of each layer, $b \in R^n$ represents the bias value of each layer, Ó represents the convolution operation, $v \in R^{k\times m\times n\times m}$ represents the weight of the convolution kernel, $c \in R^n$ represents the bias of the convolution layer, N represents the total number of convolution layers, k represents the number of the current convolution layers, m represents the dimension of the picture, and n represents the number of neurons in the convolution layer.

After obtaining the target features of the sports video frames, the feature fusion network is used to fuse the high-level features learned by the CNN, and then the fused features are fed into a convolutional network for target detection to calculate the confidence level of the corresponding category, which is calculated as shown in (2):

$$C_i^j = P_r(\text{Object}) * IoU_{\text{pred}}^{\text{truhh}}, \qquad (2)$$

where C_i^j represents the confidence level of the jth predicted box of the ith box, $P_r(\text{Object})$ represents the probability of whether the current predicted box has an object, and $IoU_{\text{pred}}^{\text{truhh}}$ represents the ratio of the intersection and concatenation of the predicted and true borders.

The coordinate position of the target frame is also calculated, and the loss function is shown in (3):

$$L_{DI\ OU} = 1 - IoU + \frac{\rho^2\left(b, b^{gt}\right)}{c^2}, \qquad (3)$$

where b and b^{gt} represent the centroids of the predicted and real boxes, respectively, $p^2(b, b^{gt})$ represents the Euclidean distance, and c represents the diagonal length of the smallest rectangle containing the two boxes.

2.2. Athlete Target Tracking. Numerous areas of study, such as digital image processing, machine learning, and artificial intelligence, focus on multiperson target tracking and identification innovation for athlete coaching. It has a broad spectrum of applications possibilities and significant research value in a range of areas, including personal interaction, video image tracking, and food information extraction. To obtain the appearance characteristics of each athlete, first, all athlete target images are extracted using the target detection algorithm, and then each athlete target image is input into a simple appearance embedding model. After that, identifying the size and position of the athlete target in the following frame, the Kalman filter method [22] is used to predict the appearance attributes of each athlete in the following frame. After detecting the size and position of the athlete targets in the next frame, the Hungarian algorithm is used to match them so that each athlete target is associated with tracking classroom athlete targets.

2.3. Real-Time Athlete Behavior Recognition. The procedure of recognizing and tracking athletic behavior is a technique of target placement and monitoring. To begin, each video clip is split into any number of frames. So each frame position's velocity condition is predicted. The most basic function is to follow the target's distance according to the form box. The overall framework of athlete behavior

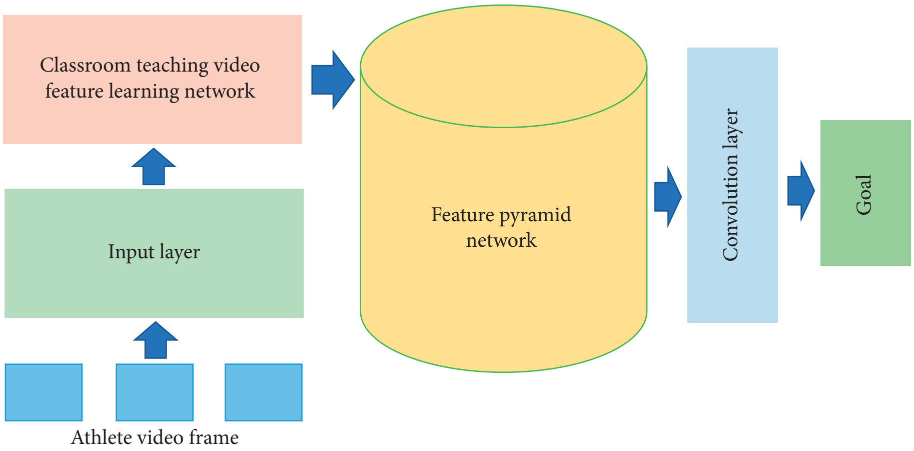

Figure 2: Flowchart of target tracking algorithm for athletes.

recognition algorithm based on deep space-time residual convolution neural network mainly includes the construction process of athlete behavior recognition model and athlete behavior recognition process [23].

All the athlete target frames are input into the athlete target tracking model to get the appearance features of the athlete targets and assign IDs to the athlete targets. After that, the collected athlete target images are preprocessed and normalized. Consequently, the spatiotemporal features of the athlete's behavior are extracted using a deep spatiotemporal residual convolutional neural network. Finally, athlete behavior recognition is achieved by classification learning. In the process of athlete behavior recognition, first, the athlete target frame is obtained by testing the classroom video stream. After that, the athlete's target frame in consecutive frames is input to the target tracking model to get the athlete target's behavior state picture stream. Finally, all of the athlete target picture streams are fed into the athlete behavior recognition model, which produces the athlete's behavior category. The specific algorithm description is shown in Algorithm 1.

Due to the characteristics of the residual structure, it can effectively extract the spatiotemporal characteristics of the athlete's behavior while reducing the amount of computation to meet the real-time demand, and the formula of the residual structure is shown in (4):

$$H(x) = f(x) + x, \tag{4}$$

where X represents the input feature, $f(x)$ represents the linear transformation, and $H(x)$ represents the output feature. In the model, cross entropy is used as the loss function, as shown in (5):

$$L = \frac{1}{N}\sum_i L_i = \frac{1}{N}\sum_i - \sum_{c=1}^{M} y_{ic}\log(p_{ic}), \tag{5}$$

where M represents the number of athlete behavior categories and y_{ic} is the indicator variable. If the category is the same as the category of sample I, it is 1; otherwise, it is 0. p_{ic} represents the prediction probability that the observed sample I belongs to category C.

3. Experimental Work and Simulations

This section introduces the experimental setup and reports on the deployment and training detailed information. Following that, we offer quantitative and qualitative examinations and also computational examination. In order to verify the effectiveness of the proposed method, the experimental window is set as a 3 × 3 local window, and the goal video of Euro 2012 and the goal video of Barcelona team in 2013-2014 seasons are selected as the input data for tactical behavior recognition.

3.1. Video Preprocessing.

3.1. Video Preprocessing. The concept "video preprocessing" refers to processes on video frames at the most basic level of understanding. In this paper, we use the input video as a video frame image as input. This can be calculated by utilizing (6):

$$x = \{x_1, x_2, \ldots, x_i, \ldots, x_N\}, \tag{6}$$

where $x_i\,(i = 1, 2, \ldots, n)$ denotes the ith frame of the xth subvideo segment and N denotes its number of frames. The input video segment is segmented by the same time interval length and expressed as

$$x = \{x_1, x_2, \ldots, x_i, \ldots, x_M\}, \tag{7}$$

where M denotes the number of segments of the subvideo segment from which the whole input video is split:

$$x = \left\{x_{j1}, x_{j2}, \ldots, x_{ji}, \ldots, x_{jM}\right\}, \quad j = 1, 2, \ldots, M. \tag{8}$$

Here, X_{jm} denotes the jth subvideo of the segmented input video, and $jm = pm, j \neq p, j, p = 1, 2, \ldots, M, q = 1, 2, \ldots, N$, and x_{jq} are the qth frame images in the jth subvideo segment.

In order to make the spatiotemporal distribution of players on the court so that they look neither too dense nor sparse [19], it is advisable to divide the subvideo segments within the range of 3 to 9. The recognition accuracy of dividing different segments is shown in Figure 3. As seen in Figure 4, the best result is obtained by dividing a video segment into 6 subvideo segments, and each subvideo segment should be less than 10 seconds.

3.2. Information Feature Extraction. The aim of feature removal is to decrease the amount of characteristics in a dataset by creating fresh features from present ones. These newly reduced sets of characteristics should be able to summarize the majority of the evidence in the novel set of characteristics.

3.2.1. Regional Division. In this work, we have used the field line detection algorithm proposed in the literature [20]. Here, the left half of the standard soccer field is used as the division object, and the right half of the field is treated in the same way. According to the geometric characteristics of the soccer field, it is divided as follows: except for the top point, three points are equidistant in the left half of the field borderline, three points are symmetrical on the bottom line, and a left half of the field containing 16 small areas can be obtained by connecting the corresponding points of the relative lines with straight lines. Figure 5 shows the schematic diagram of the soccer field after division.

3.2.2. Significant Region Discriminant Detection. For the target behavior features in the video, in order to get the regions or parts that are the interest to the human visual mechanism, this paper incorporates the concept of regional discrimination of visual saliency based on the local regression kernel to accurately extract the saliency regions within the video frames to ensure that the obtained feature information meets the observation requirements. Pixel and supervoxel saliency and video saliency map calculation are both important components of the saliency region discrimination detection process. The computation of the saliency map is given in equation (9):

$$\begin{aligned} Sal\left(p_i\right) =& N\left\{\sum_{i=1}^{L}\left[\mu\left(p_i, p_j\right)F\left(p_i, p_j\right)\right]\right\} \\ &+ \lambda N\left\{\left\{\sum_{j=1}^{L}\left[\mu\left(p_i, p_j\right)F\left(p_i, p_j\right)\right]\right\}SV\left(p_i, t\right)^{\alpha}\right\}, \end{aligned} \tag{9}$$

where $N\{\sum_{j=1}^{L}[\mu(p_i, p_j)F(p_i, p_j)]\}$ is the significance of an image space part, $\lambda N\{\{\sum_{j=1}^{L}[\mu(p_i, p_j)F(p_i, p_j)]\}SV(p_i, t)^{\alpha}\}$ is the incremental change between video frames, λ is the time increment weight, α is the time regulation factor, and n is the set of normalization operators (its value belongs to [0, 1]). After weighing the significance through the video frame significance map, each supervoxel value set is obtained, and the set is used to represent the discrimination degree of regional discrimination. The supervoxel significance is given in equation (10):

$$Sal\left(S(i)\right) = \frac{\sum_{q\in S(i)} Sal(q)}{N\left(S(i)\right)}, \tag{10}$$

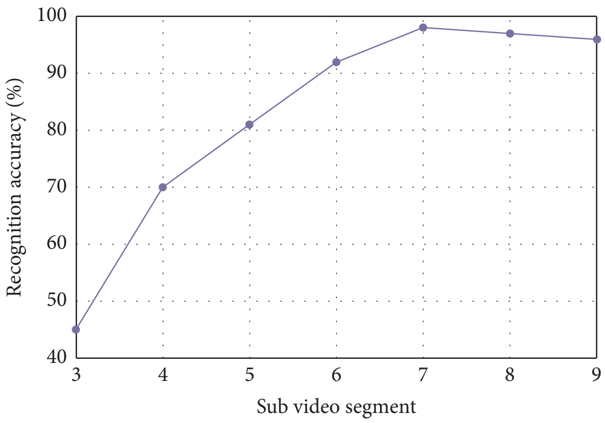

Figure 3: Split number recognition accuracy.

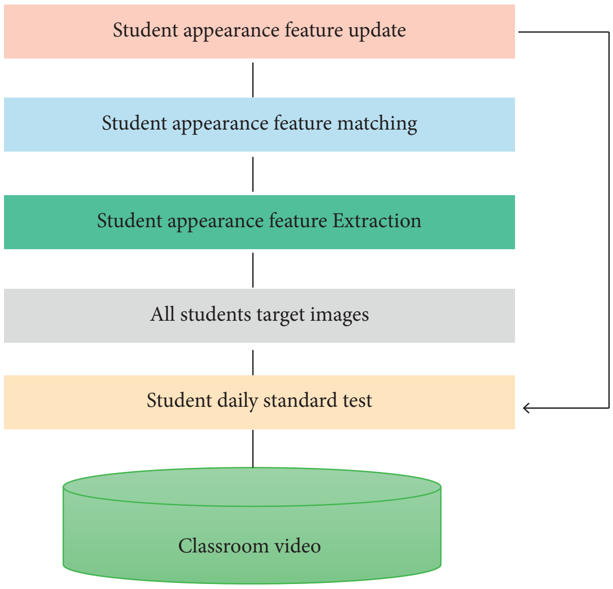
Student appearance feature update
Student appearance feature matching
Student appearance feature Extraction
All students target images
Student daily standard test
Classroom video

Figure 4: Flowchart of athlete target tracking algorithm in class.

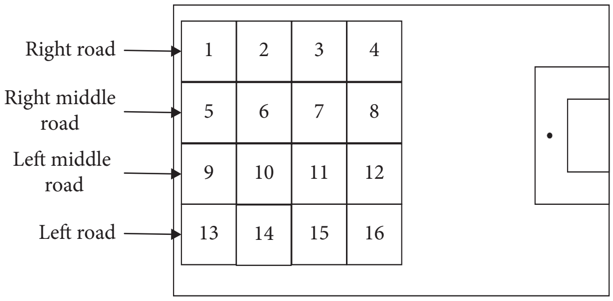

Figure 5: Schematic diagram of soccer field division.

where $S(i)$ is the ith supervoxel corresponding to the video frame, $N(S(i))$ is the number of pixels contained in the ith supervoxel, Q is the pixel point, and $Sal(q)$ is the significance of the pixel. For the discriminant significance, the significance weighted discriminant formula is

$$D\left(S(i)\right) = \begin{cases} \beta + f\left(S(i)\right), & Sal\left(S(i)\right) > S, \\ \beta \cdot Sal\left(S(i)\right) + f\left(S(i)\right), & Sal\left(S(i)\right) \leq S, \end{cases} \tag{11}$$

where $f(S(i))$ is the result after the discrimination of the ith supervoxel and β is the weight value of the balance term. When $f(S(i)) > 2$, it is regarded as the significance discrimination region of the video and the observation target region and finally achieves the regional discrimination effect [21].

3.2.3. Active Player Detection and Location. According to the three-dimensional state local spatiotemporal regression kernel proposed in this paper, an active player discrimination model directly involved in tactics is constructed. In this paper, the threshold of the local spatiotemporal regression model is set to 0.7, and the weight in the linear weighted feature fusion calculation expression is set as $A1 = 0.4$ and $A2 = 0.6$. Each activity region represents a player on the field, and if the activity is greater than or equal to the threshold, it is active; otherwise, it is inactive. Then, using the camera calibration model proposed in the document [22], create an activity map to detect the activity players of each segmented subvideo segment and locate them on the divided court area.

3.2.4. Construction and Concatenation of Feature Word Bag Model. For each subvideo segment, the obtained activity map is labeled corresponding to the location of the field segmentation area in the real spatial coordinate system; that is, let $E_j(i) = a$, $a \in \{1, 2, \ldots, 16\}$, where $E_j(i)$ represents the ith sports activity area on the subvideo segment x_i, a is the corresponding field area label, and each activity area corresponds to a football player. Due to the different colors of jerseys of different teams, the team to which the player belongs can be identified according to its color and texture information, the field areas corresponding to all active players on the subvideo segment can be labeled, and the team to which they belong can be identified.

First, all the subvideo segments divided into equal time are subjected to the same processing as described above, and then the frequency of the sports activity area of the team to be identified in each subvideo segment in each court segmentation area is calculated. The frequency is given in (10):

$$PC_{ja} = \frac{\sum_i E_j(i)}{\sum_j \sum_i E_j(i)} = a. \tag{12}$$

The feature word bag histogram λ_j of the activity area on the subvideo segment x_j is constructed, still taking the left half as an example. The horizontal axis in the histogram represents the field area label of {1, 2, . . ., 16} in the subvideo segment, and the vertical axis represents the frequency of the activity area of the experimental team in each field area. Because the subvideo segment is segmented at equal time intervals, there is a continuous relationship in time sequence. After obtaining the feature word bag histogram of each subvideo segment, the feature representation of football video tactical behavior can be obtained by concatenating its time sequence relationship $\{\lambda_1, \lambda_2, \ldots, \lambda_j, \ldots, \lambda_M\}$.

3.3. Tactical Behavior Recognition Classification. According to the position of players on the football field during the attack, three tactical behavior categories are given, namely, middle attack, side attack, and coordinated attack. The middle attack corresponds to {5, 6, 7, 8} and {9, 10, 11, 12} areas in Figure 5, and the side attack corresponds to {1, 2, 3, 4} and {13,14,15} in Figure 5, 16. In the area, active players in cooperative attacks appear evenly in the divided area; that is, the corresponding local space-time pattern is evenly distributed between the middle attack and the side attack. According to the transmission mode of the ball in the attack, it is divided into a fast counterattack and short pass penetration. Three methods classified by position are combined with two methods classified by transmission. So, six corresponding combined classification methods can be obtained. The input goal video is processed with the help of the LibSVM classifier. The intuitive classification is shown in Figure 6.

After classifying the tactical behaviors of a team or an event in the whole game according to the above methods, the main tactical behaviors used in the game can be analyzed according to the percentage of the six tactical behaviors used in the scoring process.

3.4. Case Analysis. In this paper, we apply the video of the football match to analyze all the goals scored by the Barcelona teams in the Euro 2012 match. Here, we point out the strategies that are mainly adopted by the whole match and the single team. In order to obtain recall, precision, and accuracy, the selected goal video of 2012 European Cup football match is manually marked, and the results are listed in Table 1.

After screening, there are 60 goal clips containing the six attack modes defined above. The video frames of the six attack modes in the 2012 European Cup are shown in Figure 7.

Next, take the manual marking result m as the comparison standard, and compare and analyze the computer recognition result with the manual marking result. Set the correct recognition result as h_c, the wrong recognition result as h_f, and the unrecognized result as h_m. In order to illustrate the accuracy of the recognition result, carry out multiple experiments on the input data, and then take the average value. Recall (R), precision (P), and accuracy (C) are defined in equation (13).

$$\begin{aligned} R &= \frac{h_c}{\left(h_c + h_m\right)}, \\ P &= \frac{h_c}{\left(h_c + h_f\right)}, \\ C &= \frac{h_c}{m}. \end{aligned} \tag{13}$$

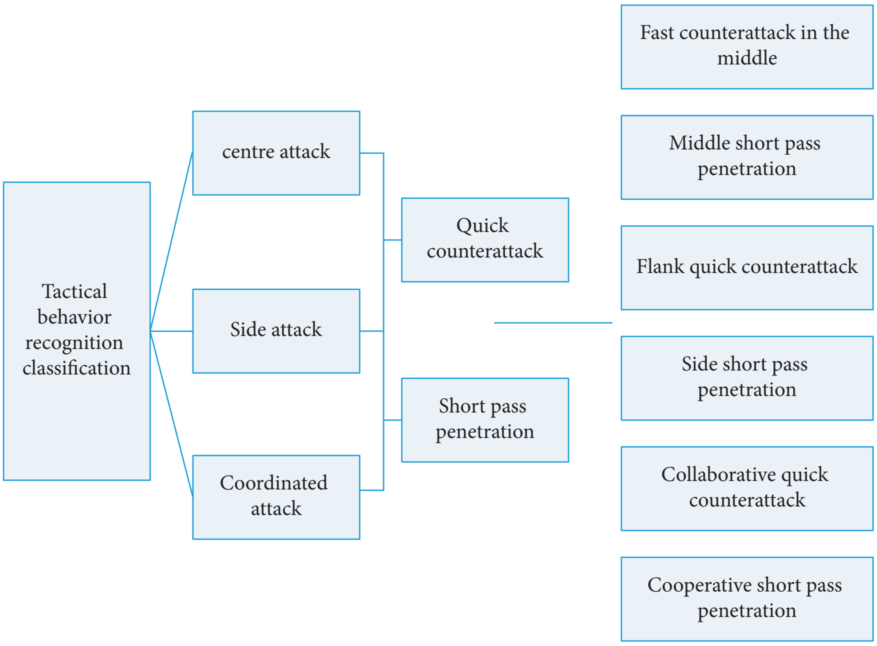
Tactical behavior recognition classification
centre attack
Side attack
Coordinated attack
Quick counterattack
Short pass penetration
Fast counterattack in the middle
Middle short pass penetration
Flank quick counterattack
Side short pass penetration
Collaborative quick counterattack
Cooperative short pass penetration

Figure 6: Identification and classification of tactical behavior on football field.

```
Input:                    Sports Video
Output:                   Behavior categories of all athletes
Initialization: Initialize target detection model $D$, target tracking model $T$, and behavior recognition model $A$
                          Initialize the athlete set image $M$<ID, $S$>, each athlete $S$ has a cache $C$ that stores the state
image and a behavior v
    For frame in video
                          boxes = $D$(frame)/ ∗ Get all athlete target images by athlete target detection ∗ /
                          IDs = $T$ (boxes)/ ∗ Update athlete image assignment IDs by athlete target tracking ∗ /
    For ID in IDs
        If ID not in $M$:
                          M.add(ID, s)/ ∗ Add the nonexistent athlete S to the athlete set $M$ ∗ /S.C.add(frame [ box
])/ ∗ Store the corresponding athlete target image to the corresponding athlete's cache $C$ ∗ /
        If S. C. size = 16: S.v = $A$(S.C)/ ∗ Identify athlete behavior by behavior recognition model ∗ /
        End if
    End for
    End for
```

Algorithm 1: Athlete behavior recognition algorithm based on deep spatiotemporal residual convolution neural network.

The identification results of the six attack modes used in the 2012 European Cup football match obtained by using this method are listed in Table 2. It can be seen from this table that the team's goals in the 2012 European Cup football match mainly adopt the middle fast counterattack, followed by the cooperative fast counterattack. Among the six tactical behaviors applied, the recognition accuracy of tactical methods of middle and side attack is higher. Furthermore, the accuracy of cooperative attacks is slightly lower than the first two, but it also reaches more than 80%. Cooperative rapid counterattack and cooperative short pass penetration have recognition accuracy of 90.6% and 80%, respectively. The recall and precision of cooperative short pass penetration, on the other hand, are 69.6% and 71.1 percent, respectively. It is lower than other tactical behavior recognition results, which is mainly due to the unclear distinction between the boundaries of the middle and side roads in the recognition process. Regarding human eye recognition, there are errors in the manual marking results, and the reason for the relatively low accuracy can be understood from the definition of cooperative attack.

Figure 8 depicts a comparison of the number and percentage of goals scored by different tactics in Euro 2012. This figure clearly shows that the percentage of 21 fast counterattacks in the middle, that is, 27.6%, is greater than the other. Furthermore, 3 side short pass penetrations have the lowest percentage at only 3.9%.

Figure 7: Video frame images of three attack modes in European Cup 2012.

Table 1: Number and percentage of goals scored by various tactics in Euro 2012.

Attack mode	Fast counterattack in the middle	Middle short pass penetration	Flank quick counterattack	Side short pass penetration	Collaborative quick counterattack	Cooperative short pass penetration	Others
Quantity	21	9	7	3	16	4	46
Percentage %	27.6	11.8	9.2	3.9	21.1	5.3	21.1

Table 2: Recognition results and accuracy rate of goal tactical behavior in European Cup 2012.

Tactical behavior	h_c	h_m	h_f	R%	P%	C%
Fast counterattack in the middle	20.1	2.2	1.2	90.1	94.4	95.7
Middle short pass penetration	8.2	1.4	1.3	85.4	86.3	91.1
Flank quick counterattack	6.7	0.5	0.4	93.1	94.4	95.7
Side short pass penetration	2.8	0.4	0.3	87.5	90.3	93.3
Collaborative quick counterattack	14.5	2.3	3.1	86.3	82.4	90.6
Cooperative short pass penetration	3.2	1.4	1.3	69.6	71.1	80.0

Figure 9 shows the comparisons of the accuracy of the six attack modes used in the 2019 [24] European Cup football match obtained during our experiment. Among the six tactical behaviors applied, the recognition accuracy of tactical methods of middle and side attack is higher. Furthermore, the accuracy of cooperative attacks is slightly lower than the first two, but it still exceeds 80%. Recognition accuracy for cooperative rapid counterattack and cooperative short pass penetration is 90.6 percent and 80 percent, respectively. The recall and precision of cooperative short pass penetration, on the other hand, are 69.6% and 71.1%, respectively.

To sum up, based on the significant detection of picture pixels based on a 2D local regression kernel, this paper proposes an activity discrimination model based on the local spatiotemporal pattern. Firstly, the activity of a central point in space-time is discriminated by calculating the feature fusion value of a central point on three orthogonal planes, and the activity map is constructed. Second, it is recommended that the feature word bag model be tailored to the actual area and that the feature histogram be combined with the color information of the research object. Finally, the feature histogram of each subvideo segment is connected in series in time order.

According to the feature selection of each behavior, the LibSVM recognition classifier is used for classification and recognition. This method effectively solves the problem of low recognition accuracy caused by incomplete trajectory extraction or too complex background due to occlusion in the trajectory-based method. This method is applied to the video of the European Cup 2012 and Spanish League football match 2013-2014 to recognize the tactical behavior of players' goals. The average accuracy is 91.3%. The experimental results verify that the proposed method has high accuracy and practicability for target behavior recognition in video.

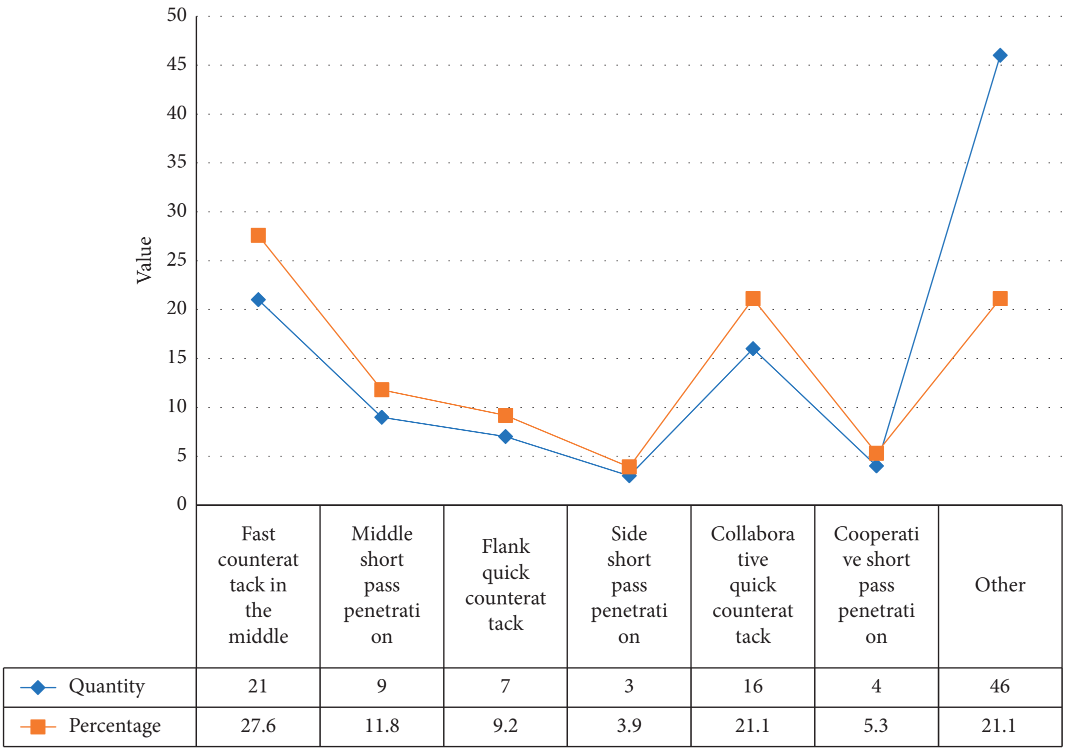

Figure 8: Comparison of the number and percentage of goals scored by different tactics in Euro 2012.

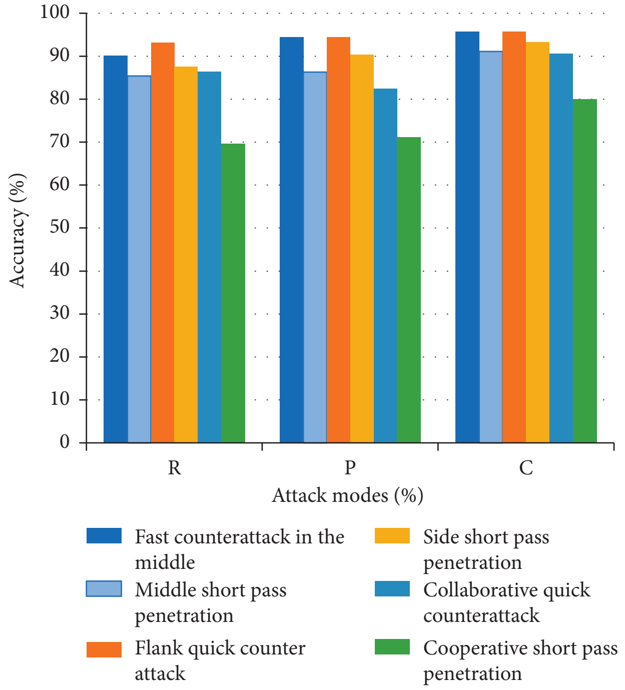

Figure 9: Comparison of accuracy rate of goal tactical behavior in European Cup 2012.

4. Conclusions

The ability to recognize and analyze sports video behavior is essential for the automatic understanding of sports. Identification and localization of action in sports video behavior are two important research topics in this context. The detection and localization of active individuals in sports video behavior are proposed in this research using a local pattern activity discrimination algorithm. This solves the problem of low recognition accuracy due to incomplete extraction of trajectories or overly complex backgrounds in the trajectory-based approach. The model is applied to the video segments of 38 matches associated with goal events in Euro 2012, and an average accuracy of 91.3% is obtained. The experimental results verify the high accuracy and practicality of the method for the recognition of target object behavior in the video. In the future, we aim to use a modified version of this paper for various sports activities such as badminton and volleyball. Also, we aim to study the proposed approach from security and privacy perspective as well.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This study was supported by the Undergraduate Teaching Engineering Construction Project, no. 201903022, and the Teaching Reform and Practice Based on the Integration of Inside and Outside Class, no. SJJG20191216.

References

[1] L. Liu, L. Shao, X. Zhen, and X. Li, "Learning discriminative key poses for action recognition," *IEEE Transactions on Cybernetics*, vol. 43, no. 6, pp. 1860–1870, 2013.

[2] C.-M. Chen and L.-H. Chen, "Novel framework for sports video analysis: a basketball case study," in *Proceedings of the 2014 IEEE International Conference on Image Processing (ICIP)*, Paris, France, January 2014.

[3] Q. Liu and Q. Liu, "Prediction of volleyball competition using machine learning and edge intelligence," *Mobile Information Systems*, vol. 2021, Article ID 5595833, 2021.

[4] D. W. Tjondronegoro and Y.-P. P. Chen, "Knowledge-discounted event detection in sports video," *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, vol. 40, no. 5, pp. 1009–1024, 2010.

[5] Z. Gao, P. Wang, H. Wang, M. Xu, and W. Li, "A review of dynamic maps for 3D human motion recognition using ConvNets and its improvement," *Neural Processing Letters*, vol. 52, no. 2, pp. 1501–1515, 2020.

[6] X.-B. Fu, S.-L. Yue, and D.-Y. Pan, "Camera-based basketball scoring detection using convolutional neural network," *International Journal of Automation and Computing*, vol. 18, no. 2, pp. 266–276, 2020.

[7] B. Li and X. Xu, "Application of artificial intelligence in basketball sport," *Journal of Education, Health and Sport*, vol. 11, no. 7, pp. 54–67, 2021.

[8] R. A. Minhas, A. Javed, A. Irtaza, M. T. Mahmood, and Y. B. Joo, "Shot classification of field sports videos using AlexNet Convolutional Neural Network," *Applied Sciences*, vol. 9, no. 3, p. 483, 2019.

[9] A. Jalal, I. Akhtar, and K. Kim, "Human posture estimation and sustainable events classification via pseudo-2d stick model and k-ary tree hashing," *Sustainability*, vol. 12, no. 23, p. 9814, 2020.

[10] B. Fasel, J. Spörri, P. Schütz, S. Lorenzetti, and K. Aminian, "An inertial sensor-based method for estimating the Athlete's Relative Joint Center positions and center of mass kinematics in alpine ski racing," *Frontiers in Physiology*, vol. 8, 2017.

[11] F. Jiang and X. Chen, "An action recognition algorithm for sprinters using machine learning," *Mobile Information Systems*, vol. 2021, pp. 1–10, Article ID 9919992, 2021.

[12] M. Rafiq, G. Rafiq, R. Agyeman, G. S. Choi, and S.-I. Jin, "Scene classification for sports video summarization using transfer learning," *Sensors*, vol. 20, no. 6, p. 1702, 2020.

[13] S. Su, J. P. Hong, J. Shi, and H. S. Park, "Predicting behaviors of basketball players from first person videos," in *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, July 2017.

[14] K. Lu, J. Chen, J. J. Little, and H. He, "Lightweight convolutional neural networks for player detection and classification," *Computer Vision and Image Understanding*, vol. 172, pp. 77–87, 2018.

[15] Y. Yang, "Research on basketball sports neural network model based on nonlinear classification," *Journal of Intelligent & Fuzzy Systems*, vol. 40, no. 4, pp. 7567–7576, 2021.

[16] N. A. Rahmad, N. A. J Sufri, N. H. Muzamil, and M. A. As'ari, "Badminton player detection using faster region convolutional neural network," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 14, no. 3, p. 1330, 2019.

[17] J. Tang, "An action recognition method for volleyball players using deep learning," *Scientific Programming*, vol. 2021, pp. 1–9, Article ID 3934443, 2021.

[18] J. Žemgulys, V. Raudonis, R. Maskeliūnas, and R. Damaševičius, "Recognition of basketball referee signals from real-time videos," *Journal of Ambient Intelligence and Humanized Computing*, vol. 11, no. 3, pp. 979–991, 2019.

[19] C. Guo, "Prediction and evaluation model of physical training for volleyball players' effect based on grey Markov theory," *Scientific Programming*, vol. 2021, Article ID 6147032, 2021.

[20] L. Wu, Z. Yang, Q. Wang et al., "Fusing motion patterns and key visual information for semantic event recognition in basketball videos," *Neurocomputing*, vol. 413, pp. 217–229, 2020.

[21] D. Cook and A. Vardy, "Towards real-time robot simulation on uneven terrain using neural networks," in *Proceedings of the 2017 International Joint Conference on Neural Networks (IJCNN)*, Anchorage, AK, USAm, May 2017.

[22] X. Fu, K. Zhang, C. Wang, and C. Fan, "Multiple player tracking in basketball court videos," *Journal of Real-Time Image Processing*, vol. 17, no. 6, pp. 1811–1828, 2020.

[23] L. Wu, Z. Yang, J. He et al., "Ontology based global and collective motion patterns for event classification in basketball videos," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 7, pp. 2178–2190, 2019.

[24] L. Chen and W. Wang, "Analysis of technical features in basketball video based on deep learning algorithm," *Signal Processing: Image Communication*, vol. 83, Article ID 115786, 2020.